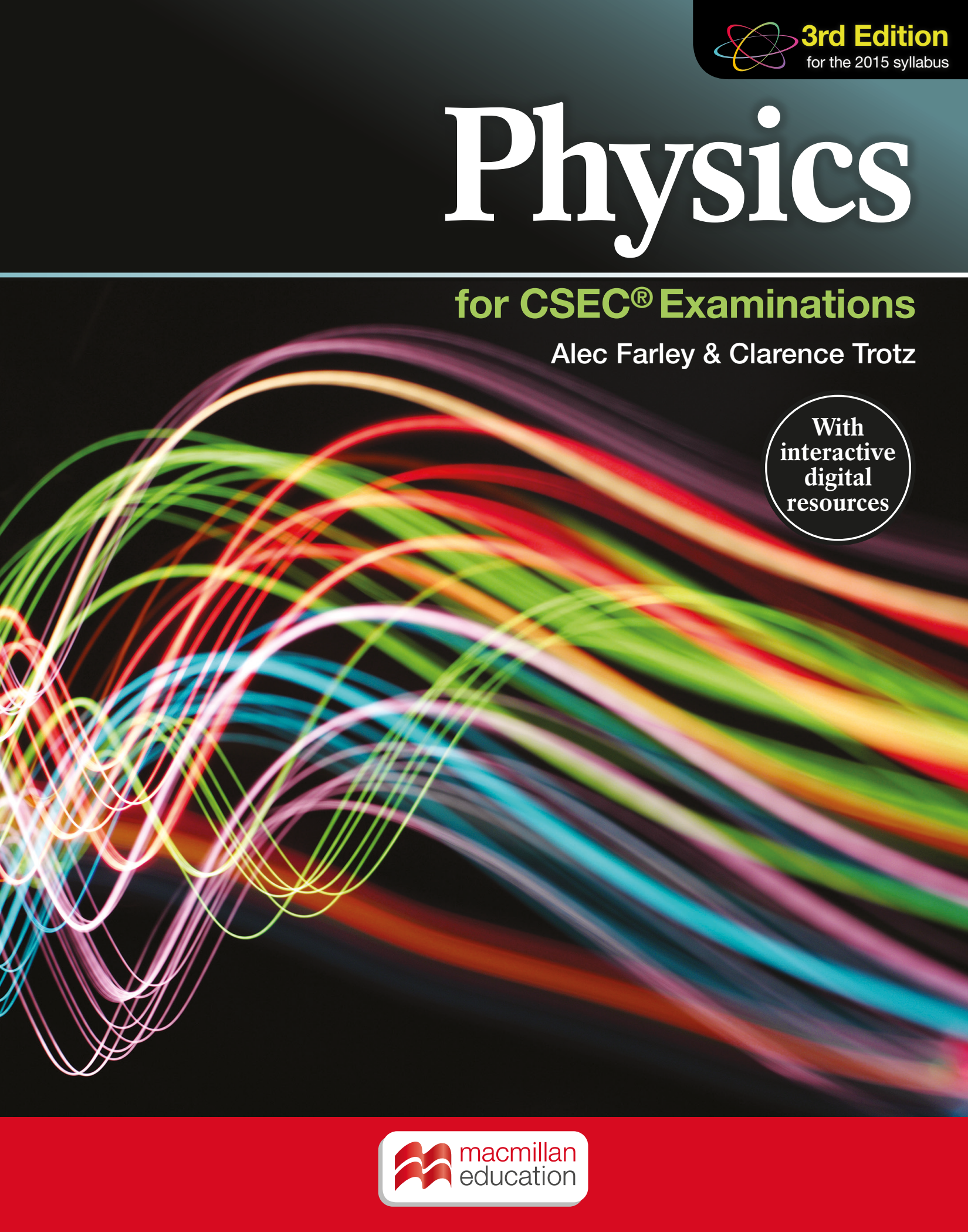

3rd Edition
for the 2015 syllabus
Physics
for CSEC® Examinations
Alec Farley & Clarence Trotz
With interactive digital resources
macmillan education

Online student resources

Log on to the Macmillan Caribbean website (www.macmillan-caribbean.com/mee) to access additional online student resources that support this Physics for CSEC® Examinations textbook. Resources include interactive questions, audio-based activities, animations and a mark book to track your progress.

Access code: **PHYS255930554769** (5 years)

Physics

for CSEC® Examinations
3rd Edition

Alec Farley and Clarence Trotz
Series Editor: Dr Mike Taylor

CSEC® is a registered trade mark of the Caribbean Examinations Council (CXC).
PHYSICS FOR CSEC® EXAMINATIONS THIRD EDITION is an independent publication and has not been authorised, sponsored, or otherwise approved by CXC.



Macmillan Education
4 Crinan Street, London, N1 9XW
A division of Macmillan Publishers Limited

Companies and representatives throughout the world.

www.macmillan-caribbean.com

ISBN 978-0-230-47947-0 AER

Text © Alec Farley & Clarence Trotz 2014
Design and illustration © Macmillan Publishers Limited 2014
The authors have asserted their rights to be identified as the authors of this work in accordance with the Copyright, Design and Patents Act 1988.

This edition published 2014
First edition published 2007

All rights reserved; no part of this publication may be reproduced, stored in a retrieval system, transmitted in any form, or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publishers.

Note to Teachers
Photocopies may be made, for classroom use, of pages 563–582 without the prior written permission of Macmillan Publishers Limited. However, please note that the copyright law, which does not normally permit multiple copying of published material, applies to the rest of this book.

Designed by Macmillan Education
Original design by Jim Weaver Design
Illustrated by Raymond Turvey (Turvey Books Ltd), TechType & Gary Ning
Cover design by Macmillan Education
Cover Credit: **Corbis**/ STScI/NASA(back cover); **Getty**/Mark Wragg (front Cover)
Picture research by Catherine Dunn
Typeset by Orchard Publishing, Cape Town, South Africa

The author and publishers would like to thank the following for permission to reproduce their photographs: **Alamy**/BSIP SA P534(cl), Alamy/Oliver Benn p196, Alamy/dbimages p450, Alamy/DBURKE p142(tl), Alamy/Andy Fletcher p176(lr), Alamy/Givaga p176, Alamy/Leslie Garland Picture Library p12(A), Alamy/David J. Green – technology pp218(tr), 16(tl), Alamy/Hero Images Inc. p280, Alamy/David Lyons p214, Alamy/Mary Evans Picture Library p386, Alamy/Trevor Mogg p546, Alamy/Purestock p98(br), Alamy/Tom Stack p452(tl), Alamy/Skyscan Photolibrary p539, Alamy/Pete Titmuss p385, Alamy/JOHN WHEELER p12(B), Alamy/A. T. Willett p171(br); **Art Directors**/Helene Rogers p5, Art Directors/Robin Spurr p371, Art Directors/Viesti Collection Inc p132; **Bureau International des Poids et Measures** p3(cl); **Corbis**/ASA/Roger Ressmeyer p114, Corbis/Ralph A. Clevenger p180, Corbis/ZAHID HUSSEIN/Reuters p376, Corbis/Reuters p P98(bc), Corbis/Erik Tham p473; **Digital Vision** p302, **Alec Farley** pp449, 468; **Getty Images** pp421, 431, Getty/Jimmy Anderson p288, Getty/Richard Cummins p399, Getty/DEA/G. DAGLI ORTI p53, Getty/Peter Dazeley p528, Getty/Daniel J. Grenier p241(c), Getty/David G Hemmings p88, Getty/Neil Holmes p142(tr), Getty/The Image Bank p520, Getty/Dr. Heinz Linke p333, Getty/Navaswan p241(a), Getty/David Rosenberg p216, Getty/Rubberball/ Mike Kemp p98(bl), Getty/Taxi p233, Getty/Mark Williamson pp p142 (cl), 463(tl); **Impact Photos**/Jorn Stjerneklar p452(tr); **Jamaica-Gleaner.com** p451(br); **Mary Evans Picture Library** P517(cr), Mary Evans/INTERFOTO/Sammlung Rauch p112; **NASA** p451(bl), p547; **Photodisc**/Getty Images pp171(bl), PhotoDisc/Getty Images \ Don Farrall (headers); **Rex Features**/Ian Hodgson /Associated Newspapers p218(bl); **Robert Harding PL** p534(tl); **Science Photo Library** p517(tl), Science Photo Library/ALEX BARTEL p416, Science Photo Library/ANDERSSON, OXFORD MOLECULAR BIOPHYSICS p291, Science Photo Library/MARTYN F. CHILLMAID p394, Science Photo Library/PHIL DAUBER p235, Science Photo Library/JAMES KING-HOLMES p483, Science Photo Library/LIBRARY OF CONGRESS p534(br), Science Photo Library/Damien Lovegrove p16(br), Science Photo Library/CORDELIA MOLLOY p419, Science Photo Library/ ALEXIS ROSENFELD p230, Science Photo Library/SOVEREIGN, ISM p533, Science Photo Library/JAMES STEVENSON p451(cl), Science Photo Library/MARK SYKES p408, Science Photo Library/US DEPARTMENT OF ENERGY p541; **Science and Society Picture Library** p519, Science and Society/Science Museum pp3(br), 4; **Mike Taylor** pp169, 184;Thinkstock p463(camera, laptop, mobile).

The author(s) and publishers are grateful for permission to reprint the following copyright material:
Extract from: 'Potential uses of terahertz radiation.' First published in the *Daily Telegraph* on 6 October 2002. © Telegraph Media Group Limited 2002. Reprinted with permission.

These materials may contain links for third party websites. We have no control over, and are not responsible for, the contents of such third party websites. Please use care when accessing them.

Although we have tried to trace and contact copyright holders before publication, in some cases this has not been possible. If contacted we will be pleased to rectify any errors or omissions at the earliest opportunity.

Contents

Series preface vii

About this book viii

Introduction: Measurement

1 Physical Quantities and Units **2**
Introduction 2
Physical quantities 2
Units for fundamental quantities 3
Multiple and sub-multiple units 4
Derived quantities 6
Units named after famous scientists 6
Chapter summary 9
Answers to ITQs 9
Examination-style questions 9

2 Precision, Accuracy and Significance **10**
Introduction 10
What is meant by 'a satisfactory result'? 11
Precision 11
Degree of significance of a result of a calculation 14
Types of instruments found in physics 14
How do errors arise? 17
Applying limits of error of measurement 18
Determining the degree of significance in the result of using a formula 20
Spreading the error 21
The meaning of a mean (or average) 22
Chapter summary 23
Answers to ITQs 24
Examination-style questions 24

3 Acquiring Experimental Skills **26**
Introduction to experimental skills 26
Experimental errors 27
Carrying out practical activities 30
Experimentation 32
Examples of possible experiments for testing practical skills 36
Criteria for the assessment of practical skills 39
Practical activity to practice and to assess manipulation and measurement (M/M) skills 41
Activity to assess planning and designing (P/D) skills 43
Issues involved in assessing analysis and interpretation (A/I) skills 47
Chapter summary 47
Answers to ITQs 48
Examination-style questions 48

1 Section A: Mechanics 51

4 Galileo Galilei and the Simple Pendulum **52**
Galileo Galilei – what was his contribution to scientific methodology? 53
What is a 'simple' pendulum? 53
Chapter summary 62
Answers to ITQs 62
Examination-style questions 63

5 Combining and Resolving Vectors **64**
Scalars and vectors 64
Resolving a vector into two components 72
Chapter summary 73
Answers to ITQs 74
Examination-style questions 74

6 Forces **75**
How various forces are produced 75
Contact forces and non-contact forces 76
Gravitational force 78
The magnitude of a gravitational force 79
Mechanical forces 81
Non-contact forces 82
Magnetic forces 83
Elastic forces 83
Nuclear forces 84
Chapter summary 84
Answers to ITQs 85
Examination-style questions 85

7 Moments and Levers **87**
Coplanar forces 87
Equilibrium 88
Finding the position of a centre of mass 95
The lever principle applied in machines 97
Stability 100
Chapter summary 101
Answers to ITQs 102
Examination-style questions 103

8 Motion in a Straight Line **104**
Galileo and Newton 104
The relationship between 'distance–time' graphs and speed 105
Dynamics and Newton's laws of motion 112
Testing Newton's laws by experiment 117
Chapter summary 122
Answers to ITQs 122
Examination-style questions 123

9 Energy, Work and Power **126**
Introduction 126
The meaning of energy 127
Types of potential energy 129
Kinetic energy 132
The principle (or law) of conservation of energy 133
Frictional force and energy 134
Power 135
Efficiency 138
Energy sources 141
Chapter summary 142
Answers to ITQs 143
Examination-style questions 144

10 Pressure and Archimedes' Principle **145**
Introduction 146
Pressure 146
Pressure due to solids 147
Pressure due to liquids at rest 149
Archimedes' principle 156
Application of Archimedes' principle to sinking and floating 159
Chapter summary 162
Answers to ITQs 163
Examination-style questions 164

Section B: Kinetic Theory and Thermal Physics 166

11 The Kinetic Model of Matter **168**
Particles of matter 168
The Kinetic Theory of matter 169
Evidence for the Kinetic Theory 169
Applying the Kinetic Theory 172
Change of state 174
Chapter summary 180
Answers to ITQs 181
Examination-style questions 182

12 Temperature and its Measurement **183**
Thermometers are everywhere 183
What is temperature? 184
Measurement of temperature 186
Thermometers in practice 190
Chapter summary 193
Answers to ITQs 194
Examination-style questions 194

13 Measuring Heat Energy **195**
The Caloric Theory of heat 196
Experiments on the nature of heat 196
Heat as energy transfer 198
Specific heat capacity 198
Specific latent heat 204
The importance of specific heat capacity and latent heat 209
Chapter summary 210
Answers to ITQs 210
Examination-style questions 211

14 Methods of Heat Transfer **213**
More ways than one! 213
Methods of heat transfer 214
Absorbers, reflectors and radiators of electromagnetic (radiant) energy 218
Practical applications of heat transfer methods 220
Chapter summary 223
Answers to ITQs 224
Examination-style questions 224

15 The Behaviour of Gases **226**
Putting gases to work 226
Boyle's law 227
Charles' law 231
The pressure law 232
The general gas equation 234
Chapter summary 235
Answers to ITQs 236
Examination-style questions 236

Section C: Waves and Light 239

16 What Are Waves and How Are They Produced? **240**
Introduction 240
How are waves produced? 241
Transverse and longitudinal waves 243
Waves in a ripple tank 245
What happens when a wave travels through a medium? 246
Matter waves and electromagnetic waves 249
Chapter summary 250
Answers to ITQs 251
Examination-style questions 251

17 Wave Characteristics **252**
Introduction 252
Frequency and period 253
Phase and wavelength 254
Displacement and amplitude 256
Wavefronts 257
Wave speed 258
Chapter summary 259
Answers to ITQs 260
Examination-style questions 261

18 Sound as Wave Motion **263**
Introduction 264
Production and propagation of sound waves 264
Reflection of sound 269
Refraction of sound 270
Interference of sound waves 273
Diffraction of sound 276
Pitch and loudness of sound 279
The audible range 279
Ultrasound 279
Chapter summary 281
Answers to ITQs 281
Examination-style questions 282

19 Electromagnetic Waves 283
Introduction 284
How are electromagnetic waves produced? 284
The nature of electromagnetic waves 284
Electromagnetic spectra 286
The complete electromagnetic spectrum 286
Special properties and uses of electromagnetic waves 288
General properties of electromagnetic radiation 291
Chapter summary 292
Answers to ITQs 293
Examination-style questions 293

20 Light 295
Introduction 296
Rival theories of light – a touch of history 296
Current theory of light: wave–particle duality 297
Use of the photon in modern technology 297
Light behaving like a wave: Young's slits 298
Light travels in straight lines 300
Eclipses of the Sun and the Moon 302
The pinhole camera 303
Chapter summary 305
Answers to ITQs 306
Examination-style questions 306

21 Reflection of Light 307
Introduction 307
The laws of reflection 308
Locating the image formed by a plane mirror 309
Features of the image formed by a plane mirror 312
Some uses of plane mirrors in physics 316
Chapter summary 316
Answers to ITQs 316
Examination-style questions 317

22 Refraction of Light and its Relation to Colour 318
Introduction 319
The laws of refraction 320
Tracing the path of light rays through rectangular blocks and triangular prisms 321
The reversibility principle 323
The refractive index: comparing refraction in transparent media 325
Finding the refractive index of a glass block by experiment 327
Critical angle for two transparent media in contact 329
Total internal reflection 330
The 90°–45°–45° prism 331
Total internal reflection in optical fibres 332
Dispersion of white light 333
Newton and colour 335
Producing a pure spectrum 326
Recombining the colours of the light spectrum 337
The colours of objects 338
Chapter summary 338
Answers to ITQs 340
Examination-style questions 340

23 Lenses 342
Lenses: what are they and how do they work? 343
The shapes of lenses 343
Lens terminology 346
How to locate (find the position of) a virtual image 354
Constructing ray diagrams 355
Construction of ray diagrams for diverging lenses 358
The differences between images produced by converging and diverging lenses 359
Solving lens problems using a scale construction 362
The lens equation (lens formula) 365
Chapter summary 367
Answers to ITQs 368
Examination-style questions 368

Section D: Electricity and Magnetism 369

24 Electrostatics 370
Static and current electricity 371
Static electric charge 371
Electron theory of charging and conduction 373
Methods of charging and discharging 375
Detecting and estimating charge 377
Distribution of charge on the surface of a conductor 378
Electric fields 379
Storing electric charge 380
Chapter summary 381
Answers to ITQs 382
Examination-style questions 382

25 Current Electricity 384
Safety first 385
A simple invention that revolutionised the world! 385
Electric current and charge 386
Alternating and direct current 387
Energy and electric circuits 388
Describing electrical circuits 390
Setting up a d.c. electrical circuit 395
Chapter summary 396
Answers to ITQs 397
Examination-style questions 397

26 More about Electrical Circuits 399
Introduction 400
Series circuits 400
Voltage (potential difference, or p.d.) in series circuits 401
Resistance in series circuits 401
Parallel circuits 403
Combined series and parallel circuits 406
Current–voltage relationships 407
Determining current–voltage relationships 408
Energy and power in an electric circuit 412
Chapter summary 412
Answers to ITQs 413
Examination-style questions 414

27 Magnetism and Electricity 416
The close association between magnetism and electricity 417
Magnetism 417
Magnetic fields around magnets 418
Magnetism from electricity 421
Domain theory of magnetism 425
Methods of making magnets 426
Chapter summary 427
Answers to ITQs 428
Examination-style questions 428

28 Electromagnetism 430
Another link between electricity and magnetism 431
Electromagnetic force 431
Applications of the electromagnetic force 434
Electromagnetic induction 436
Electromagnetic induction in a straight wire 439
Applications of electromagnetic induction 440
Chapter summary 446
Answers to ITQs 446
Examination-style questions 447

29 Electricity for Home and Society 449
Introduction: electricity at work 450
Generating electricity 451
Distributing electricity 452
Household electric circuits 453
Surges, spikes and low voltages 459
Paying for electricity 459
Portable electricity 460
Conserving electrical energy 464
Chapter summary 465
Answers to ITQs 465
Examination-style questions 466

30 Electronic Devices 468
Introduction: electronics 468
Vacuum-tube devices 469
Some uses of the oscilloscope 471
Semiconductors 473
Semiconductor 'resistors' 478
Chapter summary 479
Answers to ITQs 479
Examination-style questions 480

31 Putting Electronics to Work 483
Electronic applications 483
Switches 484
Some simple applications of electronics 485
Applications involving extensive electronic switching 491
Chapter summary 492
Answers to ITQs 493
Examination-style questions 493

Section E: The Atom 495

32 Models of The Atom 496
Early ideas 496
Dalton's model: the indestructible atom 497
Thomson's model: a cloud with light, embedded particles 497
Rutherford's model: a nuclear, planetary model 499
The Bohr model: electrons arranged in 'shells' 501
A wave model of the atom 504
Chapter summary 504
Answers to ITQs 505
Examination-style questions 505

33 Structure of the Atom 506
Particles that make up the atom 506
Stability of the atom 507
The Periodic Table 509
Bonding between atoms 511
Describing the nuclides 512
Isotopes 513
Chapter summary 513
Answers to ITQs 514
Examination-style questions 514

34 Emissions from the Nucleus 516
The discovery of radioactivity 516
Safety in handling radioactive materials 517
Detecting radioactive emissions 518
Types of emissions from the nucleus 521
Representing radioactive decay 523
Making use of emissions from the nucleus 525
Chapter summary 525
Answers to ITQs 526
Examination-style questions 526

35 Half-Life 528
Why half-life? 528
Half-life 529
Uses of radio-isotopes 532
Chapter summary 535
Answers to ITQs 536
Examination-style questions 536

36 Energy from the Nucleus 538
The energy problem of today 538
A nuclear energy alternative? 539
Energy from the nucleus 540
Obtaining energy from the nucleus 541
A nuclear fission power plant 543
Prospects for fusion reactors 545
The nuclear power debate 545
Chapter summary 547
Answers to ITQs 548
Examination-style questions 548

Section F: School-Based Assessment 551
Practical work in Physics 552
School-Based Assessment contents 563

Appendix: The Greek Alphabet 582

Index 583

Series Preface, 3rd edition

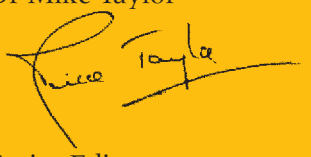
Macmillan's textbooks for the Caribbean Secondary Education Certificate (CSEC) Science subjects have been written by teachers with many years' experience of preparing students for success in their examinations. These revised third editions have been written to align with the new CXC syllabuses (to be first examined in 2015). Additional practical activities have been included to reflect the new emphasis on practical work, and new features (such as group work and discussion activities) will help teachers to cater to a variety of different learning styles within the classroom.

These books are specially designed to stimulate learning, whatever the reader's needs. Students starting a topic from scratch may need to be led through the explanation one step at a time, while those with prior knowledge of a topic may need to clarify a detail, or reinforce their understanding. Others may simply need to check that they understand the material.

Each CSEC science syllabus specifies the areas to be used for the School-Based Assessment (SBA). Each book in the series has a section designed to help students with their SBA, by offering advice on how to approach the task, presenting examples of good SBA work or suggesting suitable material to use within it. Teachers are free to photocopy these pages.

The CSEC Science series covers everything a student needs pass their CSEC examination, as well as providing a firm foundation for more advanced study at Caribbean Advanced Proficiency Examination (CAPE) level.

Dr Mike Taylor



Series Editor

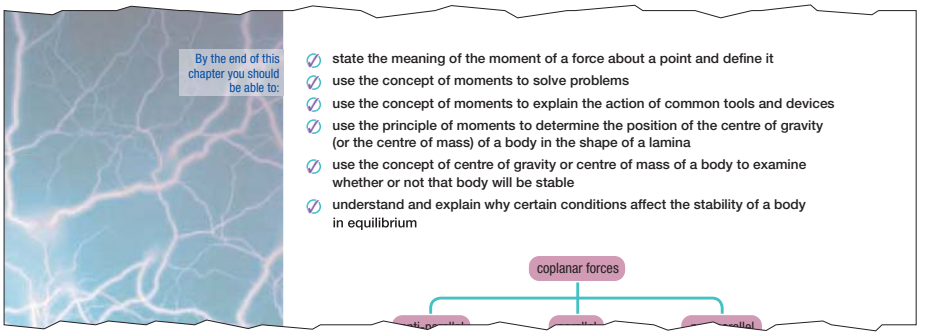
About this book

This book isn't just words on a page.

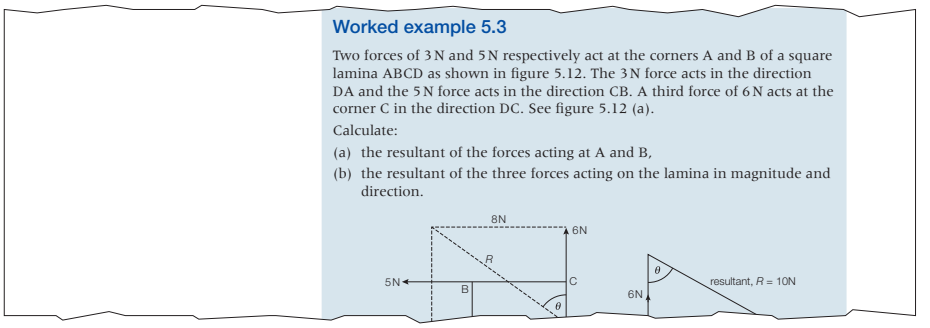
This book contains a range of different features to introduce, teach and highlight key information throughout the course. These pages explain how to use them. The larger column contains the main text and diagrams; you can read straight down it without interruption.

The smaller column contains other useful facts, so make sure you use it to check your understanding. You should remember to spend time studying the figures and diagrams as well as the text..

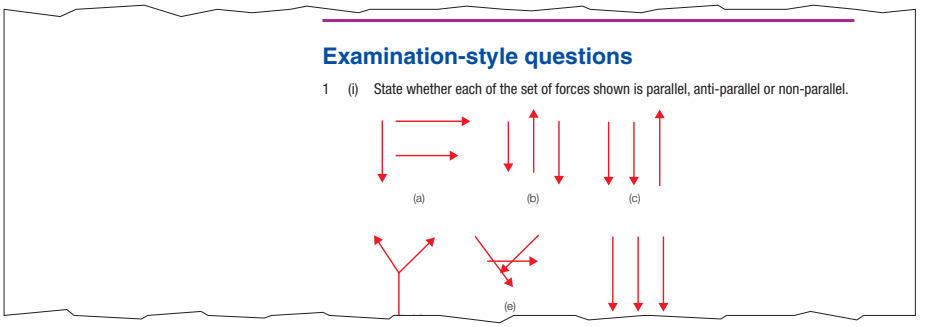
A list of objectives at the beginning of each chapter tells you what topics you will be covering. They will help you to plan and measure your learning.



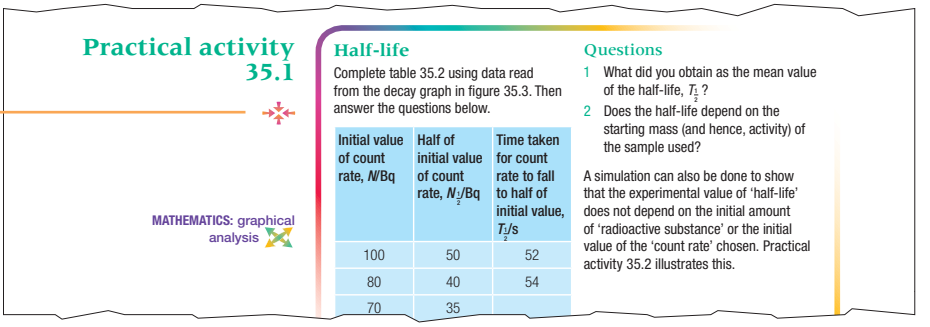
Worked examples, tables and definitions are printed in coloured boxes for easy recognition.



This is the style of question you may come across in your exam. Your teacher will suggest how you can use them, but they will measure what you have learnt and help to identify any gaps in your knowledge so you can revisit the relevant sections of the book.



This symbol denotes a group activity, where you can work with others to explore the theoretical concepts in the chapter. They will test your investigative and problem-solving skills and show real-world applications of the facts you are learning.



Where you see this icon, you will find an In-Text Question (ITQ). These are spread throughout each chapter and will help you to check your progress. If you can't answer the ITQ, you should refresh your knowledge by re-reading the relevant paragraphs in the main text.
Answers to the ITQs are found at the end of each chapter.

Activities with the 'discuss' icon give you the opportunity to show your ability to communicate scientific information, to explore and assess data, and to reach conclusions with your classmates.

Summaries of the key facts from each module will help you check your understanding.

The first time an important new word appears in the text, it is highlighted at the side. Sometimes a short definition is given beneath it, or an in-depth explanation is provided in the main text.

This symbol means that you can find additional practice for this topic on the Macmillan CSEC Science digital resources. These stand-alone components will help you to learn and revise key areas of the course. For more information please visit: http://www.macmillan-caribbean.com/pages.aspx/educationalbooks/secondary/science/interactive_science_csec/.

This icon shows how you can make links between this concept and other subjects, for example Mathematics. It is important to remember that you are not just learning facts in isolation but to think about how they relate to your world and your experiences. It also shows how different topics link within this book.

gamma decay

Gamma (γ) decay

ITQ5

By how much does the mass number of a nuclide change after (i) alpha, (ii) beta and (iii) gamma emission?

Parents and daughters

A nuclide that decays is called a parent. The nuclide that results from decay is called the daughter. Thus, in the β decay equation given, carbon-14 is the parent and nitogen-14 is the daughter. The production of an element different from the parent is called **transmutation**.

For class discussion

What would the gamma decay equation for figure 34.14 be?

Beyond $Z = 82$, nuclei tend to be very unstable on account of the very large numbers of protons. The energy required to hold the protons together in such nuclei is large. Neutrons are no longer able to hold such nuclei in a state of stability. Apart from α-particle emission, γ rays may also be given off.

Gamma emission usually occurs when a nucleus decays from an excited (high energy) state to a lower energy state. The decay of radium-226, shown in the previous equation, provides one such example. The radon-222 nucleus resulting from α decay is, at first, in a high energy state. The energy is lowered by emission of γ radiation. Figure 34.14 shows an energy level diagram depicting α emission followed by γ emission. Note that since γ radiation has no mass and no charge, the radon-222 nucleus does not change either in mass number or in atomic number following γ emission.

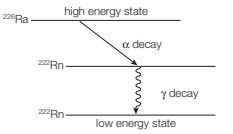


Figure 34.14 Energy level diagram showing γ emission followed by g emission.

Chapter summary

- Henri Becquerel, a French scientist, is credited with the discovery of radioactivity. Radioactivity is the giving off of emissions from the nucleus of an atom.
- Marie Curie's major accomplishments in studying radioactivity were:
  - the discovery that the amount of radiation emitted depended only on the amount of compound present;
  - the discovery of the highly radioactive elements of polonium and radium.
- Both the cloud chamber and the Geiger–Müller tube operate on the ionising effect of radiation. The cloud chamber shows tracks made by the radiation in cold air saturated with alcohol vapour. The Geiger–Müller tube is used for detecting individual particles or photons of radiation.
- Emissions from the nucleus are of three main types: alpha, beta and gamma.
- Alpha (α) particles are helium nuclei. An α particle has a charge of $+2e$ and an approximate mass of 4 u.

proton

Neutron

The word 'neutron' is derived from 'neutral', which means neither positive nor negative. Note that the names of the three types of particles all end in the suffix 'on' – neutron, proton, electron.

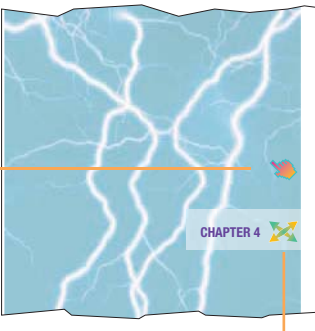
neutron

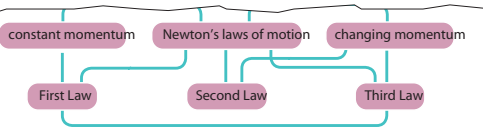
Rutherford knew that the hydrogen nucleus had the smallest charge of all known nuclei. He assumed that there was one electron orbiting the hydrogen nucleus. Since the hydrogen atom was neutral, he argued that the charge on the hydrogen nucleus must be equal and opposite to that on the electron. In 1920, Rutherford gave the name '**proton**' to the hydrogen nucleus.

Experiments also showed that the charges on other nuclei were approximately whole-number multiples of the charge on the proton. All other nuclei, Rutherford reasoned, must therefore have a whole number of protons.

On the basis of Rutherford's model, a carbon nucleus, with 6 times the charge of a hydrogen nucleus, should have 6 protons, and thus 6 times the mass. Careful experiments revealed, however, that the mass of the carbon nucleus was 12, and not 6, times that of the hydrogen nucleus.

In order to explain this 'problem', Rutherford and his colleagues assumed that, within the nucleus, there must be another kind of particle with no charge. Rutherford proposed that particles with the same mass as a proton, but with no charge, were also present in nuclei. These particles became known as '**neutrons**'. The carbon nucleus must therefore have had 6 protons and





Galileo and Newton

Motion is one of the most common experiences of everyday life, but the subject did not engage the attention of many scientists until the 16th century. Aristotle, the Greek philosopher mentioned in chapter 4, proposed that the velocity of an object depended on its speed – that a constant force applied to an object produced a constant speed in the object. He based this hypothesis on the observation that the harder the horses pulled a chariot, the faster the speed

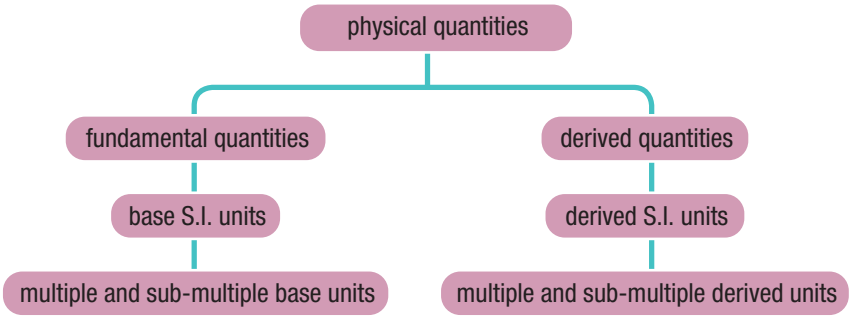
Introduction:

Measurement

1 Physical Quantities and Units

By the end of this chapter you should be able to:

- explain what is meant by a 'quantity' in physics
- state the five fundamental quantities recognised and used in physics
- explain the need for units when dealing with physical quantities
- state how the base units used in this course are defined
- explain what is meant by derived quantities and obtain their units in terms of base units
- recall and use the symbols for base units and derived units
- use multiples and sub-multiples of units
- do calculations using these multiple and sub-multiple units
- put any number into standard form



Introduction

Measurement is something we do every day to find the value or size of things.

In expressing the results we use a wide variety of units, depending on what it is that we are measuring. Whatever the result we are expressing, however, it always begins with a number and, except for very few cases, is followed by a unit. For example, a cricket score might be 85 runs, a cake recipe might mention 6 cups of flour, a salary might be 2500 dollars and the size of a hotel could be 100 rooms. The **units of measurement** in these cases are respectively runs, cups, dollars and rooms.

unit of measurement

The units we use in physics are internationally agreed, and are widely used in science, industry and technology. They are called 'S.I. units', where 'S.I.' stands for the French *Système International* ('International System'). This system of units was agreed at a conference of prominent scientists in France in 1960.

This chapter will introduce you to quantities measured in physics as well as the units in which these quantities are measured.

Physical quantities

In a school physics laboratory, there are a host of different quantities we may measure, from the length of a laboratory bench to the voltage supplied by a battery. In talking about physics, we constantly make mention of physical

quantities. These quantities and the units in which they are measured form part of the 'language of physics'.

fundamental quantity ›

In physics, seven quantities are seen as **fundamental** or basic. You will come across five of the seven in your course. These are: mass, length, time, temperature and electric current. The other two fundamental (or base) quantities are 'luminous intensity' and 'amount of substance'; but we shall not be concerned with these two in this course.

If you would like to know more about these and other concepts in the book, why not look them up online?

Internet search terms: fundamental quantities/luminous intensity/amount of substance

Each of the five fundamental quantities with which we shall be concerned is represented by a symbol, as shown in the first two columns of table 1.1.

Units for fundamental quantities

When we measure a quantity, we express the value as a number followed by a unit such as 'metre' or 'second'.

base unit ›

Each of the fundamental quantities in physics has an S.I. **base unit**. For example, the base unit of length is the metre. The base units are defined using internationally agreed standards.

The five base units most often used in physics together with their symbols are shown in the last two columns of table 1.1.

Table 1.1 Five fundamental quantities with their symbols and the corresponding S.I. base units with their symbols.

Fundamental quantity	Symbol for the quantity	Base unit	Symbol for the unit
mass	*m*	kilogram	kg
length	*l*	metre	m
time	*t*	second	s
temperature	*T*	kelvin	K
electric current	*I*	ampere	A

The units 'kelvin' and 'ampere' are named after famous scientists.

In print, the symbols for quantities in physics are given in italics as shown here – for example *T*, not T. Symbols for units are never written in the plural. For example, we would write 10 kg, not 10 kgs.

The standards kept at the International Bureau of Weights and Measures are 'primary' standards. Other 'standards', made in properly equipped laboratories and based on those at the International Bureau, are called 'secondary standards'.

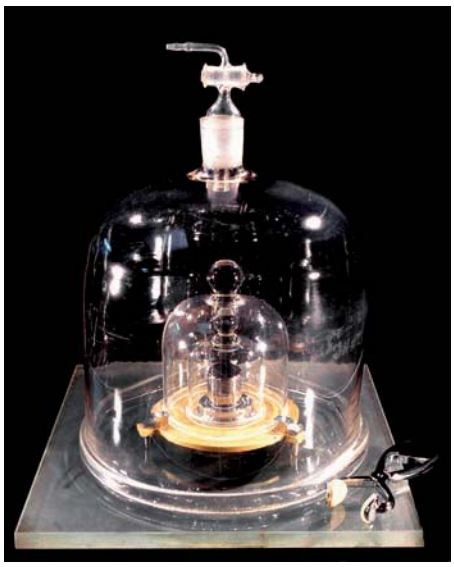


Figure 1.1 Mass is a fundamental quantity in physics. The kilogram standard mass, shown here, is kept at Sèvres, France.

The kilogram

kilogram standard ›

The kilogram is the base unit of mass. It is defined as the mass of a particular platinum-iridium cylinder kept at the International Bureau of Weights and Measures at Sèvres, near Paris, in France, and it is stored under specified conditions (figure 1.1). This cylinder is called the **kilogram standard**. All other masses are ultimately measured against this standard (figure 1.2).

Thus if we say that a certain mass is 40 kilograms, what we mean is that the mass is 40 times that of the kilogram 'standard'. Since the mass of a standard must not change with time or with change of environmental



Figure 1.2 A high-precision balance. Values for mass were ultimately based upon the primary standard kilogram at Sèvres.

conditions, the kilogram standard is made from an alloy chosen for its resistance to corrosion. It is kept under very closely controlled conditions (figure 1.1).

How might corrosion of the standard kilogram affect the measurement of mass? What are some of the consequences of a faulty standard?

The metre and the second

The values of base units must remain constant, irrespective of the environment. Because of this, the older definition of the metre, based on the separation of two fine scratches on a bar of a particular alloy, has had to be abandoned. In 1983, the metre was redefined as the distance travelled by light in a vacuum in 1/299 792 458 of a second. (You do not need to remember this number!)

The kelvin and the ampere

You don't need to know the standard definitions of the kelvin and ampere for your exam but why not look them up online?

Internet search terms: kelvin/ampere

The older definition of the second, based on the rotation of the Earth on its axis, has also been abandoned. The second was redefined in 1967, as the time for 9 192 631 770 vibrations of a particular electromagnetic wave given off by the atoms of caesium-133 (figure 1.3). (Again you do not need to remember this!)

When standardised in this way, the values of the metre and the second are not affected by environmental conditions. Further, these definitions allow the standards to be reproduced in any properly equipped laboratory anywhere in the world with the same accuracy.

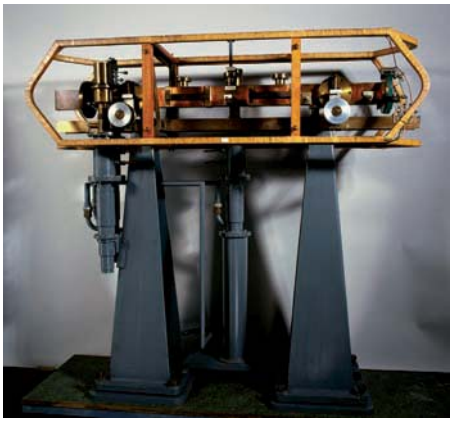


Figure 1.3 A caesium clock.

Caesium clocks are so constant that two of them will agree with each other to within 1 second in 300 000 years. This means that if the two clocks were switched on at the same time, then after they had been working for a period of 300 000 years, the times they showed would differ by no more than 1 second!

Measurements using a caesium clock show that the Earth's daily rotation is not constant, but is very gradually slowing down.

How would a change in environmental conditions affect the old metre standard?

Multiple and sub-multiple units

Imagine that two students are asked to measure the thickness of a leaf of their exercise books. One student gives the thickness as 0.2 millimetre, while the other expresses the result as 0.0002 metre. Which of these two statements gives one a better idea of the thickness?

'Sub' means 'lower than', 'less than', 'below' or 'under'.

You may have a perfectly good idea of the size of a millimetre and of the size of a metre, but it is more difficult to visualise two ten-thousandths of a metre than two-tenths of a millimetre. It is easier to visualise the size of the thickness if the student uses a rather large fraction (two-fifths) of a tiny unit, the millimetre, than it is to judge the size of a tiny fraction (two ten thousandths) of a much larger unit, the metre. There seems to be a need, therefore, for units smaller than the base unit.

sub-multiple unit ›

Such units are called **sub-multiple units**.

We also need units of length that are greater than the metre. Imagine you are to run a marathon race. Is it much easier to visualise a distance of 26 kilometres rather than 26 000 metres? Units such as the kilometre, which are larger than the fundamental unit, are called **multiple units**.

multiple unit ›

There is a need, therefore, for both multiple and sub-multiple units and which one is used will depend on the size of the quantity being measured. Whereas the sub-multiple units are fractions (which are negative powers

of ten) of the base unit, the multiple units are multiples (which are positive powers of ten) times the base unit. The factor by which the base unit is multiplied is given by a prefix, as shown in table 1.2.

Table 1.2 Prefixes for multiple and sub-multiple S.I. units.

	Prefix	Abbreviation	Power of 10
Sub-multiples	pico	p	10^{-12}
	nano	n	10^{-9}
	micro	μ	10^{-6}
	milli	m	10^{-3}
	centi	c	10^{-2}
	deci	d	10^{-1}
Multiples	deca (or deka)	da	10^{1}
	hecto	h	10^{2}
	kilo	k	10^{3}
	mega	M	10^{6}
	giga	G	10^{9}
	tera	T	10^{12}

The abbreviation 'da' is hardly ever used. Hecto- and deca- are also seldom used nowadays in physics.

A hectare is a unit of area used for land measurement. One hectare is 10 000 m^2, roughly 2.5 acres.

ITQ1
How many cm^3 are there in 1 dm^3?

micron ›

The unit 'micrometre' is sometimes called the **micron**, written as 'μ' (the Greek letter *mu*), without the m for 'metre'.

Liquid volumes in chemistry are commonly measured in dm^3.

1 decimetre^3 = 1 dm^3
= $(10^{-1}\ m)^3$
= $10^{-3}\ m^3$

This is 1 litre (l). The litre is used in chemistry and in commerce (figure 1.4).



Figure 1.4 A filling station in Trinidad. What is the unit used on the pump for measuring the quantity of gasoline bought?

ITQ2
Express the following numbers in standard form:
(i) 2000
(ii) 0.000 001 000
(iii) 700×10^{-9}
(iv) 1000×10^{-8}
(v) 123 456.789

Standard form

This is a convenient way of writing very large numbers or very small fractions, by expressing them as a number between 1 and 9.9999… multiplied by an appropriate power of 10.

It would be impractical to write a very small fraction or a very large number with a very large number of zeros. For example, we certainly would not write the Avogadro Number as 602 000… with 21 zeros following the digit 2. We therefore use standard form to represent this huge number.

Standard form is also referred to as scientific notation. Examples are:

ITQ3
Express, in standard form:
(i) 10 000 milliseconds in seconds
(ii) 2000 km in metres
(iii) 6400 km in megametres
(iv) 0.002 g in micrograms
Take care to write the symbols correctly.

- 6.02×10^{23} (the Avogadro Number);
- $3.00 \times 10^{8}\,\text{m s}^{-1}$ (velocity of light in air);
- $1.00 \times 10^{-9}\,\text{m}$ (one nanometre, the unit of wavelength of electromagnetic radiation).

Worked example 1.1

Express the very small fraction 0.000 042 in standard form.

Solution

The first factor of the required number must begin with the digits 4 and 2. From the rule, the first factor must be 4.2 (a number between 1 and 9.9999…). Since to obtain 4.2 we must multiply the given fraction by 10^5, in order to preserve the value of the decimal fraction, we must now multiply 4.2 by 10^{-5}.

So $0.000\,042 = (0.000\,042 \times 10^{+5}) \times 10^{-5} = 4.2 \times 10^{-5}$

MATHEMATICS: indices 

Derived quantities

derived quantity ›

Fundamental quantities can be multiplied or divided by other fundamental quantities to give **derived quantities**. For example, length (as in distance travelled) may be divided by time to find a speed. The resulting quantity, speed in this example, is called a derived quantity. A derived quantity may also be multiplied or divided by either a fundamental quantity or another derived quantity to give a different derived quantity. Examples are

$\frac{\text{force}}{\text{area}}$ = pressure

$\frac{\text{mass}}{\text{volume}}$ = density

density ›

$$\text{density} = \frac{\text{mass}}{\text{volume}}$$

Some other examples of derived S.I. quantities are shown in table 1.3.

Table 1.3 Some derived S.I. quantities.

Derived quantity	Unit	Symbol for unit	Derivation
acceleration	metre per second squared	m s^{-2}	
area	metre squared	m^2	
density	kilogram per metre cubed	kg m^{-3}	
electric charge	coulomb	C	$1\,\text{C} = 1\,\text{A s}$
energy	joule	J	$1\,\text{J} = 1\,\text{N m}$
force	newton	N	$1\,\text{N} = 1\,\text{kg m s}^{-2}$
momentum	kilogram metre per second	kg m s^{-1}	
potential difference	volt	V	$1\,\text{V} = 1\,\text{J C}^{-1}$
power	watt	W	$1\,\text{W} = 1\,\text{J s}^{-1}$
pressure	pascal	Pa	$1\,\text{Pa} = 1\,\text{N m}^{-2} = 1\,\text{kg m}^{-1}\,\text{s}^{-2}$
velocity	metre per second	m s^{-1}	
volume	metre cubed	m^3	

Note that there is no dot (.) or dash (-) between the base units or their symbols.

So 1 C = 1 A s and *not* 1 A.s or 1 A-s; and $5\,\text{m s}^{-1}$, not $5\,\text{m.s.}^{-1}$.

Another reason for leaving a space between the m and the s in the second example, so we write m s^{-1}, not ms^{-1}, is because 'ms' means 'millisecond'.

derived unit ›

The units used to measure derived quantities are called **derived units**. For example,

$$\text{derived quantity, speed} = \frac{\text{distance travelled (length)}}{\text{time taken (time)}}$$

$$\text{derived unit of speed} = \frac{\text{metre}}{\text{second}} = \frac{\text{m}}{\text{s}} = \text{m s}^{-1}$$

So the unit of speed is the derived unit, m/s (or m s^{-1}). We say that the unit of speed has the dimensions 'metre/second' or (m s^{-1}).

Any unit obtained by multiplying or dividing base units is a derived unit.

Units named after famous scientists

The unit of force has the dimensions mass × (length/time^2), and its unit can also be shown as kg m s^{-2}.

newton ›

Units are often named after scientists who have made a significant contribution to a particular field of study. For example, Isaac Newton did a lot of work in the area of mechanics, which is mostly about the effect of forces, and so the unit of force, the **newton**, has been named after him. The symbol for this unit is N.

At one time the symbol for the unit of volume 'the litre' was an upper case 'L'. Can you find out whether there was ever any famous scientist named 'Litre'?

pascal ›

The rule about using units named after people is this. If the name of the unit is written in full, the first letter of the unit name is written in lower case. For example, we might write 10 newtons, or 6 joules. If, however, a symbol is used to represent the unit, then the first letter of the name is written in upper case, for example 10 N, or 6 J. To state a pressure using its unit '**pascal**' (called after the French mathematician Blaise Pascal) we might write 100 Pa or 100 pascals. We do not use simply P to represent pascal, since P has been used to represent a unit of a different physical quantity.

Worked example 1.2

We are using milligrams (mg) rather than kilograms (kg) in order to avoid having to deal with a negative power of 10 (250 mg = 250×10^{-3} g).

A large stock bottle contains a number of medicinal tablets, each having a mass of 250 mg. The mass of all the tablets is 0.5 kg. Calculate the number of tablets in the bottle.

Solution

To find the number of tablets in the bottle, we must divide the total mass of all the tablets by the mass of one tablet. Before we do the division, however, we must have both masses *in the same unit* (milligrams, for convenience).

$$\begin{aligned} 0.5\,\text{kg} &= 0.5 \times 10^3\,\text{g} \\ &= 0.5 \times 10^3 \times 10^3\,\text{mg (since 1 g = 1000 mg)} \\ &= 0.5 \times 10^6\,\text{mg} \end{aligned}$$

So

$$\begin{aligned} \text{number of tablets} &= \frac{0.5 \times 10^6\,\text{mg}}{250\,\text{mg}} \\ &= 2 \times 10^3 \\ &= 2000 \end{aligned}$$

Note: Whenever we divide a quantity by another quantity of the same nature, provided that the units of the quantity are exactly the same, they will 'cancel out' and we are left with a number without a unit.

Such a number is said to be 'dimensionless'. It is also described as a 'pure number'. So here, 2000 is a pure number.

ITQ4

The relative density, R.D., of a substance is defined as:

R.D. = (density of the substance)/(density of water).

Calculate the R.D. of aluminium whose density is 2700 kg m^{-3}, the density of water being 1000 kg m^{-3}. Is relative density a dimensionless quantity?

Worked example 1.3

A piece of cork has a mass of 10 g and a volume of 40 cm³. Calculate its density.

Solution

We have density $= \frac{\text{mass}}{\text{volume}}$

So density of cork $= \frac{10\,\text{g}}{40\,\text{cm}^3}$

$= 0.25\,\text{g cm}^{-3}$ (or $0.25\,\text{g/cm}^3$)

Note: As you can see, g divided by cm³ gives g cm⁻³. This way of writing the unit is called 'index notation'. Units may also be written using the 'solidus' (/), for example g/cm³, but index notation is preferred for your Physics examination.

First we quote the formula that will give a solution. Then we substitute in the formula using the units for each quantity (g and cm³). This helps to make clear the unit of the result.

ITQ5

A young walker does a 15 km race in 1.5 hours. Calculate her average speed in (i) m s⁻¹, and (ii) m min⁻¹.

Practical activity 1.1

S.I. units and standard form

Recommended: Chose four teams of two students each from within the class.

- Activity (i) could be a competition among the four teams.
- Activity (ii) could be a speed competition among four different teams where a certain amount of time is allowed for the four teams to supply the answers to each part.
- Activity (iii) could be for all students in the class.

The following four S.I. units were named after famous scientists:

watt joule pascal kelvin

(i) Arrange a presentation to be delivered by four teams drawn from your class on the following topics (ask your teacher to be moderator):

  (a) the area of physics to which each of these scientists made a significant contribution and what were the contributions made by each;

  (b) the quantity measured in each of the four units, with examples.

(ii) Write down, using figures and unit abbreviations:

  (a) sixty-four thousand watts;

  (b) three thousandths of a milliwatt;

  (c) four hundred pascals;

  (d) one hundred thousand kilopascals;

  (e) three thousandths of an ampere;

  (f) fifty nanoamperes.

(iii) Express each of the values (a)–(f) in (ii) above in standard form.

Chapter summary

- Both fundamental quantities and derived quantities are used in physics.
- Five of the seven fundamental quantities are mass, length, time, temperature and electric current.
- Derived quantities are obtained by multiplying and/or dividing two or more fundamental quantities.
- The base unit is that size of the quantity with which other quantities of the same kind are compared. The base units in the S.I. system are as follows: kilogram for mass; metre for length; second for time; kelvin for temperature; and ampere for electric current.
- A quantity symbol is normally represented in print by a letter in *italic* type; a unit abbreviation is always represented by a letter in ordinary (upright or roman) type.
- Whenever a unit named after a scientist is written out in full, the first letter is always in lower case, but the first letter of the symbol for the unit is an upper-case letter.
- A multiple or a sub-multiple unit is sometimes used to express the size of a quantity in order to make it easier to visualise the size of that quantity.
- Three base units named after scientists are the kelvin (K), the ampere (A) and the pascal (Pa).

Answers to ITQs

ITQ1 $1000\,\text{cm}^3$

ITQ2 **(i)** 2.000×10^{3} ; **(ii)** 1.000×10^{-6} ; **(iii)** 7.00×10^{-7} ; **(iv)** 1.000×10^{-5}; **(v)** $1.234\,567\,89 \times 10^{5}$

ITQ3 **(i)** 1.000×10^{0}; **(ii)** $2.000 \times 10^{6}\,\text{m}$; **(iii)** $6.400\,\text{Mm}$; **(iv)** $2.00 \times 10^{3}\,\mu\text{g}$

ITQ4 2.700; Yes

ITQ5 $2.8\,\text{m s}^{-1}$; $167\,\text{m min}^{-1}$

Examination-style questions

1 State whether each of the following statements is true or false.

(i) There are seven fundamental quantities in physics.

(ii) One of these fundamental quantities is electric current.

(iii) Three other fundamental quantities are mass, length and time.

(iv) If kilogram is added to kilogram, we have a derived unit.

(v) Speed is defined as (distance covered)/(time taken), and so the unit of speed, m s^{-1}, is a derived unit.

(vi) The S.I. system of units was adopted in 1960 at a conference in England.

2 Which of the following is the best way to express the dimensions of a rectangular laboratory bench (i.e. a large desk surface)?

A $1.04\,\text{m} \times 400\,\text{cm}$

B $1.04\,\text{m} \times 4.00\,\text{m}$

C $104\,\text{cm} \times 4.00\,\text{m}$

D $1040\,\text{mm} \times 4000\,\text{mm}$

3 The surface area of the laboratory bench in question 2 above is 4.16 square metres. Explain why this area is called a derived quantity.

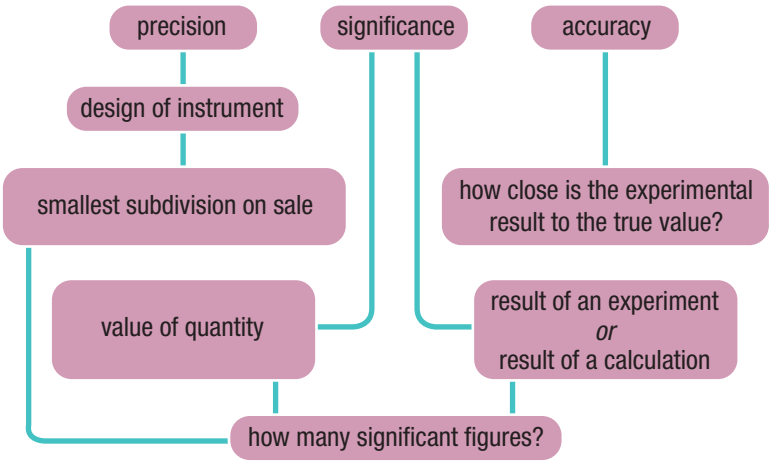
4 Why do you think the platinum-iridium metre standard was abandoned and a new definition adopted for the metre?

5 How was the second defined before the atomic (caesium) clock was developed?

2 Precision, Accuracy and Significance

By the end of this chapter you should be able to:

- distinguish between precision and accuracy
- state what is meant by precision and by the degree of precision of a measuring instrument
- state what is meant by the uncertainty or error of a reading
- explain how to determine the degree of precision of measuring instruments commonly found in the laboratory, both analogue and digital
- understand what is meant by the term 'significant figures'
- determine the degree of significance of a recorded reading
- express the result of an experiment with the correct degree of significance and the result of a numerical problem to a justifiable number of significant figures
- explain what is meant by a mean and apply this concept



Introduction

In science we constantly use instruments to measure quantities of many different kinds and the quantities we measure cover a very wide range of values. Whenever we measure a quantity we are keen to know how near our result is to the 'correct' value or, as it is properly called, the 'accepted' value. How near our value obtained by experiment is to the accepted value of the quantity is a question of its accuracy and since any reading we take with an instrument will be subject to uncertainty or 'error', we have to agree that no reading we take will be exactly the 'true' value.

Although we would like to know the true value, we never can, since any instrument we use will read with a certain degree of uncertainty depending on the sophistication of its design; the greater this is, the closer to the 'true' value we will get, always assuming that the instrument is correctly used. This chapter is concerned with this aspect of measurement – uncertainty or error – and this will lead to a consideration of significance, a specific objective of your syllabus. The question of significance and significant figures is not decided arbitrarily. It

is based on a consideration of uncertainty of measurement, which, in turn, is based on what is called the 'precision' of instruments used.

In arriving at the final result for an experiment that we carried out, we would have used values obtained for different quantities related to the determination. Each of these values would have been subject to uncertainty and this would have depended on the instrument used to measure it. If the uncertainty of all the instruments used was very small, then the uncertainty of the overall result would also have been very small and this result would have been a satisfactory one. If, however, even one of the instruments used was of high uncertainty or low precision, then our final result would also have been of high uncertainty and low precision. One inaccurate instrument can spoil our whole set of results.

When doing practical work we constantly measure and as we measure we should always be conscious of the presence of error or uncertainty at every step.

What is meant by 'a satisfactory result'?

A satisfactory result for the value of a quantity is one that is near to the 'true' value or the 'accepted' or 'expected' value of that quantity. How do you know whether or not a value is near to the 'true' value if *no one* knows the true value? The answer is that by correctly using sophisticated instruments that offer a high degree of precision, and by employing the correct theoretical procedure, we can obtain a pretty good idea of the narrow range within which this (true) value lies. The narrower the range within which the value lies, the better is the result of the experiment or investigation – the nearer the result will be to the 'true' value. The aim of the experimenter should always be to obtain the 'best' possible result with the equipment provided. This would suggest, therefore, that the range within which a result lies depends, in part, on the degree of sophistication (or **degree of precision**) of the apparatus used.

degree of precision ›

A satisfactory result for an experiment, therefore, is one which lies within a narrow range of possible values, or one which has very small 'limits of error'.

Precision

A precise instrument?

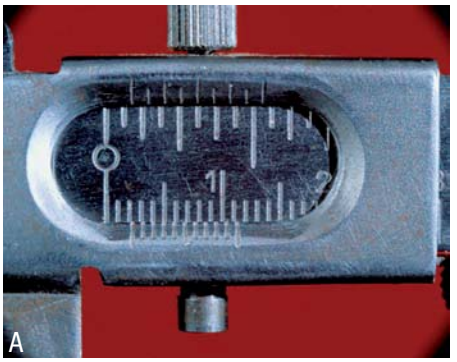
- What is meant by the term 'precise instrument?
- What is it that makes one instrument more precise than another one that measures the same quantity?
- How precise is the instrument we are using?
- Are we making the best use of what the instrument provides in order to obtain the best value for the quantity?

In other words, what we want to know is how capable the instrument is, when properly used. How capable is it of giving a reading that is as near as possible to the 'true' value of the quantity being measured? This is what we mean by the term '**precision**'. How near is the measured value to the 'true' value?

precision ›

A precise instrument is one that is so well designed that the value obtained for a quantity measured with it lies within a very narrow range of possible values, so that the size of the uncertainty or the error is small. If another instrument designed to determine the value of the same quantity can produce results which lie in a narrower range, this instrument offers better precision.

A good demonstration of this is to compare the micrometer screw gauge with the vernier caliper (figure 2.1).





B

Figure 2.1 Instrument A, the vernier caliper, is not as precise as instrument B, the micrometer screw gauge, since whereas A will measure with a precision of 0.1 mm, B will measure with a precision of 0.01 mm, one-tenth of 0.1 mm. So the micrometer gauge is 10 times as precise as the vernier caliper, but the caliper is 10 times as precise as a metre rule calibrated in millimetres.

precision ›

degree of significance ›

For a very good account of the structure and use of the screw gauge see http://www.tutorvista.com/content/physics/physics-i/measurement-and-experimentation/screw-gauge.php.

For an account of the structure and use of the vernier caliper see http://www.upscale.utoronto.ca/PVB/Harrison/Vernier/Vernier.html or http://www.technologystudent.com/equip1/vernier3.htm.

Carefully examine the structure of these two instruments. Whereas the micrometer gauge has a maximum range of uncertainty (or error) of 0.01 mm, the caliper has a maximum uncertainty or error of 0.1 mm. Although inferior to the screw gauge, the vernier caliper is superior to the simple metre rule, whose uncertainty of 1 mm is ten times larger than that of the vernier caliper.

The micrometer screw gauge may be used to obtain quite precise measurements of very small values of length, for example: wire diameters, the thickness of thin sheets, the diameter of small spheres like ball-bearings and the thickness of sheets of paper. The caliper may be used for measuring larger thicknesses and lengths up to about 8 cm.

Range of uncertainty of the result of an experiment

How can we know the range of uncertainty of the result of an experiment or an investigation that was based on a number of measurements of different quantities? The answer is that we could combine uncertainties of different readings and so arrive at a value together with its uncertainty. At this level, however, we will not do this, but use a 'rule of thumb' to arrive at an overall figure corrected to show a certain degree of significance or expressed to a certain number of significant figures. As said before, however, we do not do this arbitrarily, but use the precision or significance of values involved in the experiment.

Degree of precision of an instrument

The precision of an instrument is the maximum error there can be in measuring the size of a quantity with that instrument. Precision is also related to the width of the range of values within which the true value lies. We have seen that the micrometer screw gauge is a more precise instrument than the vernier caliper.

When we measure a value with an instrument, the result should tell the reader something about the size of the smallest subdivision on the scale of that instrument or, put another way, it should say something about the **precision** of the instrument used. Whether or not this is the case will be indicated by the **degree of significance** shown by the result.

For example, the external diameter of glass tubing may be stated as 0.56 cm if it was measured with a vernier caliper. This tells the reader that the caliper measured to the nearest 0.01 cm, as would be expected. If a micrometer screw gauge was used, however, the value declared might be 0.557 cm, which shows a precision of 0.001 cm, better than that of the caliper. So the place value of the final digit should tell us the precision of the instrument that was used and possibly what was the instrument used to measure the value. So the degree of significance shown in a result should reflect the degree of precision of the instrument used to obtain it. It is clear, then, that we should be careful when expressing results of experimentation or calculation to a certain number of significant figures, since this has important implications and will be interpreted in a certain (scientific) way.

The place value of the final digit also tells the reader the maximum error or the maximum uncertainty of the value, since this is the same as the precision of the instrument. The '6' of 0.56 cm above has a place value of one hundredth of a centimetre and this 0.01 cm is the maximum error in the value as measured with the caliper. In the same way, the maximum error in the micrometer reading of the wire diameter was 0.001 cm. This also represents the precision of the micrometer. When we speak of maximum error, we mean

the maximum difference between the 'true' value and the value obtained from the measurement.

Limits of error of a reading

When we express a value to a certain number of significant figures the place value of the final digit must tell the reader the maximum error in the value being represented. This will enable the reader to determine the limits within which the true value lies, or the limits of error, of the reading. Here is an example.

By expressing my height as 165 cm (correct to 3 significant figures) I am telling the reader that the maximum error in this value is 1 cm. The final digit of this value, 5, has a place value of unity (1) and so 1 cm must be the maximum error in this value. This is my height to the nearest (unit) centimetre.

My height therefore lies somewhere between 164.5 cm and 165.5 cm.

The difference between these two values is 1 cm. The range of uncertainty is 1 cm and the limits of error are +0.5 cm and –0.5 cm, written as ± 0.5 cm. We may say then that:

$$164.5\,\text{cm} \leq \text{my height} < 165.5\,\text{cm}$$

or $\text{my height} = (165 \pm 0.5)\,\text{cm}$

If, however, my height was 167 cm and I was asked to express it to 2 significant figures, I would have to say 170 cm.

Here the digits 1 and 7 would be significant, but not the 0. The 0 must be used, however, since without it the value (17) would be quite wrong.

The place value of the second of the two significant figures, 7, may be used to determine the limits within which the height lies as follows:

The place value of the digit 7, the second significant figure, is 10, and therefore the maximum error in this value is 10 cm. This means that my height could be anywhere between 165 cm (on the low side) and 175 cm (on the high side).

In fact my height would have been 167 cm, which is within the range calculated. We could also describe this procedure as expressing my height to the nearest 10 cm. Here again, the maximum error in the height is roughly 10, the place value of the second digit of the two significant ones. In the first case I might have measured my height with a tape graduated at 1 cm intervals, but in the second, I might have used a tape with intervals of 10 cm.

Estimating a value on a uniform (or linear) scale

The precision or uncertainty of an instrument with a uniform scale is usually taken to be the value of the smallest subdivision of the scale of that instrument and in using such a scale you would generally read to the nearest one-half of a subdivision. So if you used a metre rule graduated in millimetres to measure the diameter of the tubing considered above, you would have to judge the second figure after the 0.5 and would normally read it to be a 5 and hardly a 6, since it is not easy to distinguish between 0.5 and 0.6 of a millimetre with the naked eye.

The advantage of the vernier caliper is that it allows an easy determination of the second decimal place. Similarly the advantage of using a screw gauge over a caliper is that it allows a pretty accurate determination of the third decimal place in centimetres. It would be impossible to do this with the naked eye.

The difference between the precisions of these two instruments is due to the difference in the features offered by the two devices. Whereas the screw gauge

The precision of an instrument is the maximum error there can be in measuring the size of a quantity with that instrument. It is normally taken to be the value of the smallest subdivision of the scale.

would allow a reading to the nearest 0.001 cm, the vernier caliper will give a reading to only 0.01 cm. We could say that the screw gauge is ten times more precise than the vernier caliper.

Degree of significance of a result of a calculation

When we have obtained a set of readings, using different instruments, each with its own degree of precision, we often combine the readings to give an overall result for the value of a quantity. In declaring the result we would be expected to indicate the uncertainty in that result by using an appropriate degree of significance. This degree of significance will tell the reader what is the possible range within which the true value of the quantity lies.

Worked example 2.1

A student was asked to measure the length and breadth of a rectangle and from these measurements to find the area of the rectangle. He found the mean dimensions of the rectangle (perhaps measured with a vernier caliper) to be 6.02 cm × 4.15 cm. How should he express the area of the rectangle to an appropriate number of significant figures?

Solution

Area of the rectangle $= l \times b$

$= 6.02\,\text{cm} \times 4.15\,\text{cm}$

$= 24.983\,\text{cm}^2$

This is called the 'raw' (uncorrected) value of the area.

Rule of thumb for significance

The theoretical rule (which we will not use) suggests that we should express the area to the nearest tenth (0.1) cm. The answer should therefore be given as 25.0 cm^2.

To determine how many significant figures should be used, rather than using the strict mathematical rule for finding the overall uncertainty, we will use a 'rule of thumb' to decide the issue.

The 'rule of thumb' says that the significance must be no greater than the significance of the least significant (here the less significant, since there are only two) of the quantities used in the calculation. There is a three-figure significance in both cases, and so we would use the same degree of significance, three figures, in declaring our answer for the area.

This rule gives the same result, 25.0 cm^2, as shown in the wing note.

Types of instruments found in physics

Instruments found in physics fall into two main groups: analogue and digital. Both of these types have their advantages and disadvantages. Digital instruments are currently more common and are certainly easier to use.

Analogue instruments with uniform or linear scales

graduation ›
calibrated ›
analogue device ›

Before discussing precision in detail, we should note that some of the instruments we use have scales with equally spaced markings (or **graduations**) along them and others have a window on which the reading appears. Graduated (or **calibrated**) scales with markings are found in all **analogue** instruments (figures 2.2 and 2.3). In digital instruments, however, there is no calibrated scale, but a blank display window or read-out window on which the value appears. Table 2.1 gives examples of some analogue instruments. Figure 2.5 shows a **digital** thermometer.

digital device ›

Table 2.1 Analogue instruments with graduated scales.

Quantity	Measuring instruments
mass	spring balance, lever balance
length	tape measure, metre rule, vernier caliper, micrometer screw gauge
time	clock, stop-clock or stopwatch (of course, digital versions of these are also available)
volume	burette, measuring cylinder
temperature	mercury thermometer
force	newton-meter, spring balance

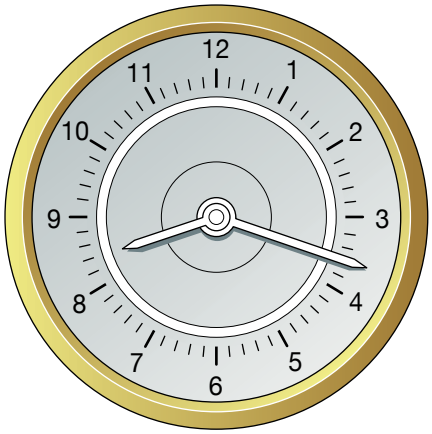


Figure 2.2 An analogue clock has a graduated scale.

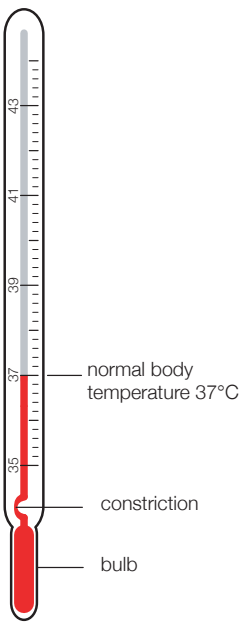


Figure 2.3 An analogue clinical thermometer. It will give a correct indication for the temperature being measured, but it is not easy to read accurately. What reading is it showing?

Analogue devices

An analogue device will give an indication for any value of the quantity which it measures within a given range. This means that the pointer on the scale can be anywhere between the limits of the scale. Whether the correct value of the quantity can be read, however, is quite a different matter. No one will be able to determine the exact value indicated by the pointer of the scale. An experimenter will be able to read correctly up to a point, but the rest of the value will be very much in doubt. This is where the uncertainty comes in. This is one disadvantage of the analogue scale. How great this uncertainty is will depend on the degree of sophistication of the instrument.

The accuracy and precision of readings given by analogue scales are always subject to the features of the instrument and it depends to some extent on the skill and experience of the experimenter in reading scales.

Digital devices

Digital devices may accept an analogue signal of some kind, but need to convert this analogue signal into discrete computer 'bits' in order to process the information. It is for this reason that a digital device, unlike an analogue one, cannot indicate on its display window the exact value of a quantity; it must first convert the information to discrete 'bits' (thereby changing the value somewhat), before displaying the value of it. This is why readings given by such instruments are given in multiples of a basic 'bit'. Thus if a 'bit' of voltage produced by a digital voltmeter is 0.002 volts, all voltages read by that voltmeter will be multiples of that bit and all readings given by the instrument will end in a 0, 2, 4, 6 or 8. A reading of 1.6 volts will mean that the meter is putting out 800 bits. The disadvantage of the digital device, therefore, is that, unlike an analogue device, it cannot give an 'output' of any value of the

quantity it is measuring, How close the output is depends on the size of the bit: the smaller the 'bit', the closer the reading to the actual value and the greater the precision of the digital instrument.

Precision of digital instruments with a display window and no scale



Figure 2.4 This top-pan balance reads 'correct to 1 gram'. It is digital and gives its reading in a display window. Its maximum error or uncertainty is 1 gram.

In figure 2.4 (a digital top-pan balance) the balance which weighs 'correct to 1 gram', has 'bits' of information each of size 1 gram. It is as if the balance can only deal with whole bits of mass of 1 gram each. The balance can therefore register only the mass of whole numbers of bits of 1 gram. We cannot have fractions of a gram. The precision of the balance is 1 g. One gram is therefore the maximum uncertainty or error in the reading for the mass of the object. Suppose, for example, that an object is placed on the balance and the reading '140 g' appears in the display window. Then the value of the mass of this object may lie anywhere between 139.5 g and 140.5 g. If m is the mass of the object, then

$$139.5\,\text{g} \leq m < 140.5\,\text{g}$$

This means that the mass is somewhere in the range (139.5–140.5 g).

The maximum uncertainty in the reading for the mass is therefore 1 g, and this is the degree of precision of the balance. We may write the value of the mass as

$$m = 140\,\text{g} \pm 0.5\,\text{g}$$

Note that 140 is half-way between the limits of 139.5 and 140.5.

limits of error ›
limits of uncertainty ›

The '**limits of error**' or the 'limits of uncertainty' of the reading for the mass are ± 0.5 g (read 'plus or minus 0.5 g'). The **limits of uncertainty** of a reading are plus or minus one-half of the degree of precision of the instrument.

If a digital balance had a precision of 10 g, then the limits of error would be ± 5 g. Such a balance would read masses to the nearest 10 g. Thus if a mass of 143 g was placed on this balance it would show a mass of 140 g, the 1 and the 4 being the significant figures but not the 0 at the end. Note that the final digit of the two which are significant (1 and 4) is the 4, which tells us that the mass is measured to the nearest 10 g, since the place value of the digit 4 is ten.



Figure 2.5 A digital thermometer. What is the range of values within which the true value lies?

In instruments with a digital readout, all the digits of the reading are shown in the display window (figure 2.5). There is no need for the observer to use his judgement in determining a reading. The uncertainty in the measurement does not depend on the skill or experience of the experimenter; it is a feature of the instrument alone. The last digit of a reading given by a digital instrument depends largely on the precision of that instrument.

ITQ1

What are the limits of uncertainty of the reading given by a digital balance which shows a reading of 6.7 g? What is the range within which the mass of this object lies? Can the last digit of a reading of a mass given by this balance be any digit from 0 to 9?

ITQ2

If you were told that a certain balance reads 'correct to the nearest 5 grams', why would you not expect it to give a reading of 267 g for the mass of an object? What would you expect the balance to read for a mass of 267 g?

Precision of instruments with a graduated scale

uniform scale | linear scale
non-uniform | non-linear

A graduated scale is one that has markings or graduations. If the markings are equally spaced, the scale is called a **uniform scale** (or a **linear scale**); if not, the scale is **non-uniform** (or **non-linear**) (figure 2.6).

Graduated scales are marked with divisions and subdivisions. The value represented by the smallest subdivision is regarded as the precision of the instrument. So, on a metre rule graduated in millimetres, we say that the precision is 1 mm; and for a thermometer whose smallest subdivision represents a change of 0.5°C, we say that the precision is 0.5°C. The limits of error in readings taken with these two instruments are then ±0.5 mm for the metre rule and ±0.25°C for the thermometer.

The same applies to other instruments with a graduated scale such as the burette, the measuring cylinder, the newton-meter, the ammeter and the voltmeter, all of which usually have uniformly graduated scales. For each of these, the maximum error in a reading will be the value of the smallest subdivision along the scale. Readings obtained with the instrument should be expressed to the nearest one-half of one scale division.

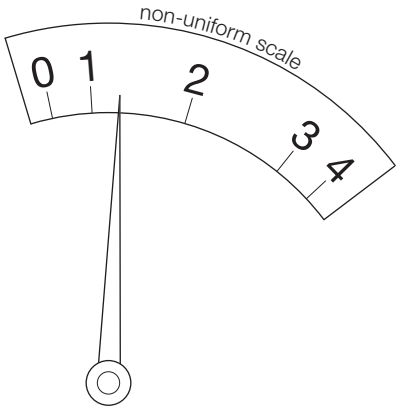


Figure 2.6 The non-linear scale of a moving-iron ammeter. Although each subdivision represents 1 A, the subdivisions are not the same size.

The precision of other instruments with linear scales

As we have already seen, the precision of an instrument with a linear scale is often taken to be the value of the smallest subdivision on the scale of the instrument. This rule works very well for most of the instruments you will use in your physics course. Table 2.2 gives some examples.

Table 2.2 Precision of some instruments with linear scales.

Instrument	Number of subdivisions in each division	Degree of precision
burette	10 per cm^3	0.1 cm^3
thermometer	2 per °C	0.5°C
newton-meter	10 per 5 newton	0.5 N
stopwatch (conventional)	5 ticks every 2 seconds	0.4 s
stop-clock	2 ticks every second	0.5 s

The precision of an instrument is a characteristic of that particular instrument; it depends mainly on its design.

How do errors arise?

Reaction error

How important is a reaction time error of 0.2 s in 'starting off the blocks' in a flat race? This would be very important if the race was over a short distance such as 100 m or 200 m. It would not be very important over long distances

The measurement of a length or a volume is the measurement of a static quantity. However, a time interval begins at a particular moment and ends at another. Unless the stopwatch is started and stopped at these precise moments, the timing of the length of the interval will have an error. Such an error has nothing to do with the instrument. It is called a 'reaction' error and it depends on the quickness of the reflexes of the user. Errors like this are often much larger than the maximum uncertainty (precision) of the stopwatch. We will see later in this chapter how the effect of reaction errors may be reduced.

Parallax, random and zero errors

Other types of errors that you will come across in your experimental work are parallax errors (figure 2.7), random errors and zero errors. Like reaction error, the first two of these are caused by the human limitations of the experimenter, but the last is due to a defect in a measuring instrument. These errors, and how to reduce their effects, will be discussed in chapter 3.

CHAPTER 3

ITQ3

Why is it best to read analogue scales with only one eye and not both eyes?

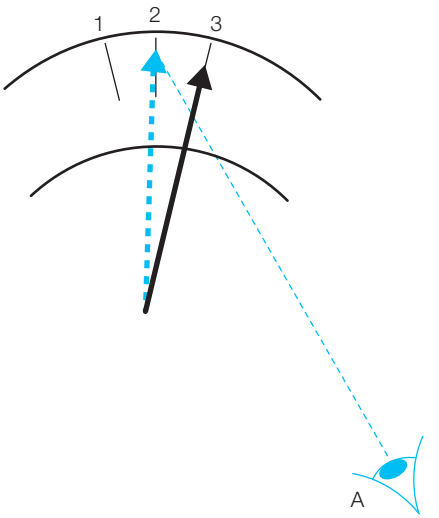


Figure 2.7 A parallax error. Looked at with one eye along a vertical line passing through the pointer, the pointer reads 3 (the correct reading). Looked at from a position slightly to the right of the correct one, the reading will appear to be smaller, how much smaller depending on how far away from the vertical the line of sight of the eye is.

Applying limits of error of measurement

Sums and differences

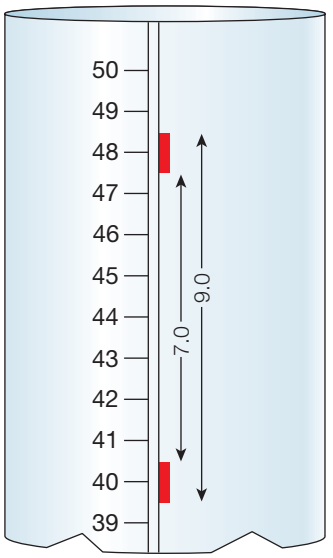


Figure 2.8 Error of measurement.

I want to measure the volume of a stone by placing it into a measuring cylinder partly filled with water. The cylinder has a precision of 1 cm^3. (It is graduated in 1 cm^3 divisions.) Before the stone is put in the cylinder, the reading for the water level is 40 ± 0.5 cm^3. With the stone in the water, the reading is 48 ± 0.5 cm^3.

We now show that the result for the volume of the stone, taking these limits into account, could be anywhere between 7.0 cm^3 and 9.0 cm^3 – a range of 2 cm^3 (figure 2.8).

The calculated value is (48 ± 0.5) cm^3 – (40 ± 0.5) cm^3, or 8 ± 1 cm^3. The largest possible difference is (8 +1) cm^3 or 9 cm^3 and the least is (8 – 1) cm^3 or 7 cm^3. Clearly, the maximum error is 2 cm^3. As a percentage of the calculated value, the largest possible error is (2/8) × 100. This is 25% – a huge percentage error! To reduce this percentage error we could:

- use a more precise instrument (this reduces the numerator in the percentage error calculation) – the uncertainty will be less;
- use a larger stone in order to have a larger divisor.

It would not be justifiable to give the result for the volume of the stone as 8.0 cm^3, since this would suggest that the result is correct to the nearest 0.1 cm^3 with a percentage error of a little over 1%. It would be more appropriate to give the volume as 8 cm^3, suggesting that the percentage error is nearer to 1 part in 8 or 12%. This error is surely nearer to the estimated maximum error of 25% than would be 1 part in 80 or about 1%.

A useful rule to remember is:

Try to avoid measuring a small quantity as the difference of two quantities, since the maximum error in the small difference will be the sum of the individual errors in the quantities being subtracted. This will lead to a large percentage error in the small difference between the masses.

ITQ4

What would the percentage uncertainty have been if we had used the same stone and a cylinder with a precision of 0.5 cm³?

This is useful advice, since you may well find it necessary to measure small masses in executing your SBA.

The effect of calculating errors in a sum is less severe since, although the errors are added as for differences, the overall error will be found as a percentage of a sum which will always be larger than a difference.

You will often find the term 'significant figures' shortened to 'sig. fig.' or 'sig. figs'.

Significant figures

One way of expressing the result of a measurement or a calculation to show the degree of precision or uncertainty is to express it with a meaningful number of significant figures. The last digit of this value will give the reader an idea of the range within which the true value lies.

Degree of significance and significant figures in the result of a calculation

When we have obtained a set of readings, using different instruments each with its own degree of precision, we often combine the readings to give an overall result for the value of a quantity. We must understand and remember that the degree of significance used in declaring a result is based on the degree of precision and is not arbitrary.

Worked example 2.2

A current of 5.05 amperes is passed for 1200 s in a mass of 100 g of crushed ice at 0°C. The supply voltage is 12 V. What mass of ice will be melted in that time? The specific latent heat of fusion of ice, $l = 3.34 \times 10^5\,\text{J kg}^{-1}$.

Solution

The theory that will lead to the solution is that the electrical energy transformed to heat = the heat absorbed to melt m kg of ice at 0°C to water at 0°C.

So $V \times I \times t = m \times l$, assuming no heat loss

where m = mass of ice which melts and l = specific latent heat of fusion of ice.

Giving $m = VIt/l$

$= (12\,\text{V} \times 5.05\,\text{A} \times 1200\,\text{s})/3.34 \times 10^5\,\text{J kg}^{-1}$

$= 21772.45... \times 10^{-5}\,\text{kg}$

$= 2.1772... \times 10^{-1}\,\text{kg}$

(change to standard form before applying significance method)

We have expressed the result in standard form in order to facilitate the process leading to the final 'corrected' result. The question now arises as to how many significant figures to use in submitting the answer. This number is not decided arbitrarily. We use the 'rule of thumb'. We look to see which is the least significant of the quantities given in the data and we find that the voltage

value is the least significant. It has the least number of significant figures, only two. We therefore express the result to two significant figures.

The mass of ice that melts will therefore be 2.2×10^{-1} kg or 220 g.

One use of standard form or scientific notation

An easy way to express a number to a certain number of significant figures is first to write the number in standard form and then to express the first factor to the required number of significant figures.

Worked example 2.3

Express 317.112 to (a) 1 sig. fig.; (b) 2 sig. fig.; (c) 3 sig. fig.

Solution

First we express the number in standard form: $317.112 = 3.171\ 12 \times 10^2$. The power of 10 has nothing to do with the process, it is the first factor which needs our attention.

(a) To 1 sig. fig. the value is 3×10^2 (maximum uncertainty = 1×10^2).

(b) To 2 sig. fig. the value is 3.2×10^2 (maximum uncertainty = $0.1 \times 10^2 = 10$).

(c) To 3 sig. fig. the value is 3.17×10^2 (maximum uncertainty = $0.01 \times 10^2 = 1$).

The number of significant figures used gives some idea of the maximum error in the value.

Determining the degree of significance in the result of using a formula

Consider the following worked example.

Worked example 2.4

Calculate the area of a circle of radius: (i) 6 cm; (ii) 6.0 cm; (iii) 6.00 cm. In each case give your answer to an appropriate number of significant figures. Assume $\pi = 3.142$.

Solution

The area of the circle, $A = \pi \times r^2$

$= 3.142 \times 6^2\ \text{cm}^2$

$= 113.112\ \text{cm}^2$ (by calculator)

$= 1.13112 \times 10^2$

In (i) the degree of significance of the radius is 1 place, and that of π is 4, and so the degree of significance of the area should also be that of the radius, or 1 place. Remember that the significance of the result is taken to be that of the lesser of the two significances involved, namely 1 place. The area of the circle should therefore be given to 1 sig. fig.

The area = $1 \times 10\ \text{cm}^2$

In (ii) the radius is 6.0 cm and the area will be given to 2 sig. fig., so the area = $1.1\ \text{cm}^2 \times 10^2\ \text{cm}^2$.

In (iii) the radius is 6.00 cm and the area will be given to 3 sig. fig. The area will be $1.13 \times 10^2\ \text{cm}^2$.

ITQ5

In each of the following quantities the number shown in brackets is the number of significant figures to which the number has been expressed: (i) 2340 (3); (ii) 2.34 (3); (iii) 0.2340 (4); (iv) 002 34 (1); (v) 100 (2).

(a) In which of the quantities could this not be the case?

(b) In each of the cases which are possible what is the maximum uncertainty (or error) in the quantity?

(c) In which of the possible cases is the uncertainty:

- **greatest?**
- **least?**

ITQ6

The distance from the centre of the earth to the centre of the moon is quoted by one source as 384 403 km. Express this distance:

- **correct to the nearest 100 km**
- **correct to 4 significant figures.**

Which of these two values is more precise? Justify your answer.

What would be the range within which the distance lies in the more precise value?

This solution has been done in accordance with the 'rule of thumb' mentioned earlier. We have taken the significance of the least significant value as our guide in determining the significance of the answer.

Spreading the error

The following worked example highlights what is referred to as 'spreading the error' of a moderately precise instrument.

Worked example 2.5

A student is given a metre rule and 10 ball-bearings and asked to find the average diameter of one of the ball-bearings. He places the ball-bearings in contact with one another in a groove between two wooden blocks and measures the value of 10 diameters. He finds that 10 diameters take up a length of 5.1 cm. Determine the average diameter of one of the ball-bearings with the appropriate degree of significance and with the limits of uncertainty.

Solution

The total length of the 10 ball-bearings is measured as 5.1 cm (significance = 0.1 cm or 1 mm for a metre rule. Since the length is given to the nearest 0.1 cm, we may take the maximum error or uncertainty in this length to be 0.1 cm or 1 mm. We may therefore express the total diameter of 10 ball-bearings as 5.1 ± 0.05 cm.

If 10 diameters measure (5.1 ± 0.05) cm, then 1 diameter will measure $\frac{(5.1 \pm 0.05)\text{ cm}}{10}$, i.e. 0.51 ± 0.005 cm.

ITQ7

If only one ball-bearing was measured, what might its diameter be, to an appropriate number of significant figures?

ITQ8

The thickness of 100 leaves of a book was found to be 14.2 mm. Calculate the thickness of one page: (i) to an appropriate number of significant figures; (ii) with its limits of error.

ITQ9

Sometimes in the building trades and in industry an article is produced with a quoted 'degree of tolerance', which is normally written as a percentage. Thus, the value of a resistor may be given as R = 10.0 ohm ± 2%.

(i) Find out what 'tolerance' means in this case.

(ii) Is the degree of tolerance an indication of precision?

(iii) Would you say that a produce with a very low tolerance has been produced very precisely? Justify your answer.

The error is only ±0.005 cm because the one measurement contained the 10 individual errors added together. We describe this procedure as 'spreading the error', since the modest 1 mm of uncertainty offered by the metre rule has been 'spread' over ten ball bearings, thereby giving to each a smaller uncertainty or error.

We would do something similar if we were:

- finding the thickness of a sheet of paper, or
- measuring the length of someone's stride, or
- measuring the period of a simple pendulum, or
- even measuring the volume or the mass of a drop of liquid.

Worked example 2.5 shows that we can measure the value of a small quantity to a high degree of precision even with an instrument that is only moderately precise. We measure the total value for a large number of these quantities taken together and 'spread the error or the uncertainty' over the number of objects being measured.

In the case of stopwatches, where the precision can be as good as 0.01 s, we must consider the reaction error of the experimenter to be more important than the precision error of the instrument when we use it to time short time intervals. Since the reaction error of the person might be as much as 0.2 s, or even more, some 20 times as large as the uncertainty of the watch, we would ignore the uncertainty of the watch, the less important of the two uncertainties, and take into account only the reaction error, which is by far the more important of the two.

The meaning of a mean (or average)

mean ›

When we find the mean of a quantity, we are finding a value that is representative of a number of values of that quantity. When we find the **mean** age of a class of students, we are calculating an age that can be used to represent all the ages in that particular class. Here we are finding a mean of a set of separate values which are not related to one another. We might, on the other hand, find a mean value for a single quantity like the diameter of a piece of glass tubing, which could have slightly different values when measured at different places along the tube. In both cases the mean value would be representative of one 'thing' – the class in one case and the diameter of the tubing in the other.

There might, for instance, be slight variation in the external diameter of a piece of glass tubing from place to place along its length. In this case and in the one before, the mean is a representative value of a quantity, the ages in the class or the diameter of the glass tubing.

With the class of students, we have a population of fixed size and so we have a fixed number of ages to add together and divide by the number of students in the class. Our result will therefore be a mean of one fixed value. In the case of the glass tubing, however, we could add together any number of diameters and then divide the sum by that number to find the mean. In the first case we would obtain a mean of definite value. In the second, the value obtained would depend on how many values of the diameter we use and where along the tube the diameter is measured. We could, therefore, in this latter case obtain a number of different values for the mean.

It will be clear that the larger the number of diameter measurements taken, the better and the more representative will be the mean value that we find.

Worked example 2.6

A student used a digital stopwatch, reading to the nearest 0.1 s, to measure the time for ten oscillations of a simple pendulum. The student, however, has a reaction time of 0.2 s. She measures the time for ten swings as 20.4 s.

(i) What is the maximum percentage uncertainty in this value?

(ii) Calculate the maximum percentage error in the time for one swing.

(iii) Calculate the mean time for one swing to an appropriate number of sig. figs.

What would have been the maximum percentage uncertainty if the student had taken the time for 20 swings?

Solution

Since the student was just as likely to start timing too early as she was to start too late, we must add the two uncertainties of 0.2 s.

(i) Time for 10 swings $= 20.4\,\text{s} \pm 0.4\,\text{s}$
$= 20.4\,\text{s} \pm 2\%$ (since $0.4 / 20.4 = 0.02 = 2\%$)

(ii) Time for 1 swing $= (20.4\,\text{s} \pm 0.4\,\text{s}) / 10 = 2.04\,\text{s} \pm 0.04\,\text{s}$
$= 2.04 \pm 2\%$

Note that although the maximum percentage uncertainty in the time for one swing is the same, i.e. 2%, the actual error is only ±0.04 s. Had the student measured the time for only one swing, the error would have been as much as 0.4 s, ten times as large (spreading the error!).

(iii) The maximum error in one swing = 2 × 0.04 s = 0.08 s, which we will call 0.1 s.

Note carefully

(i) The starting time is 0 s ± 0.2 s, and the stopping time is 20.4 s ± 0.2 s. So the time interval = (20.4 – 0) ± 0.4 s = (20.4 ± 0.04) s.

(ii) As the number of swings timed increases by a given factor, the uncertainty in the time for one swing reduces by the same factor. By using a large number of swings, we are 'spreading' the uncertainty, and the effect of the uncertainty on each swing is correspondingly reduced.

(iii) If the student had used 20 swings, she would most probably have obtained a time of about 2 × 20.4 s. This time would, however, have had the same uncertainty of ± 0.4 s. The maximum percentage uncertainty in the time for one swing would therefore have been

$$\frac{40.8 \times \pm 0.4\,\text{s}}{20} = 2.04 \pm 0.02\,\text{s}$$

This gives an uncertainty of ±1%, half as much as before. By increasing the number of swings by a factor of 2, we have reduced the uncertainty in the value of one swing by ½.

It therefore is the best to time as many swings as is practicable.

The oscillations of a simple pendulum will be studied in some depth in the next chapter.

CHAPTER 3

The degree of significance must therefore reflect this fact and so we must express the time for one swing to the nearest tenth (0.1) of a second and not to the nearest 0.01 s.

So the time for one swing should be stated as 2.0 s (nearest tenth of a second) since 2.04 corrected to the nearest tenth is 2.0.

Chapter summary

- We judge the accuracy of the result obtained for the value of a quantity by the closeness or otherwise of the result to the 'true' value.
- We judge the precision of the result obtained for the value of a quantity by the width of the range of values within which the true value lies.
- The precision of an instrument is the maximum error that is likely to be made when taking a reading with that instrument.
- The degree of precision of an analogue instrument is generally taken as the value of the smallest subdivision on the scale of the instrument.
- The precision of a digital instrument is the value to which the instrument 'reads correct'. It is the maximum error possible in a reading given by that instrument.
- The overall actual error or uncertainty in a sum or a difference of a set of quantities is the sum of the actual errors in the individual quantities. Errors are always added, never subtracted.
- Generally, in a number expressed to a certain degree of significance, the last digit gives the maximum uncertainty of the result.
- A useful guiding principle (or a 'rule of thumb') in deciding how many significant figures to use is that the degree of significance of the result should be no greater than that of the least precise of the values used in the calculation. Therefore you should use the same number of significant figures as that in the least precise of the values used in the calculation.
- The 'rule of thumb' is applied as follows:
  1. Examine the values in the numerical expression to see which of them is the least precise. The least precise value is the one with the least number of significant figures. (It is showing the largest fractional or percentage error.)
  2. Express the final result to the same number of significant figures as that contained by the least precise of the measurements used.
- In recording a reading given by an instrument, we must use a degree of significance that reflects the degree of precision of that instrument. This degree of significance is shown by the number of significant figures used to represent the value.
- The mean value of a property possessed by an object is the value that may be taken to represent that property for the whole object.
- If we are to obtain a mean value of a quantity that is near to the 'true' value, we must calculate it from a number of readings of that quantity that is as large as is practicable.
- When carrying out practical work, we should take care when measuring those quantities in which the percentage uncertainty is greatest.

Answers to ITQs

ITQ1 The limits of uncertainty are ±0.05 g; range 6.65–6.75 g; yes.
ITQ2 Because any mass it measures must be given as a multiple of 5. So the reading will be 265 g.
ITQ3 To be sure that there is only one line of sight which is vertical and passes through the pointer and on to the correct reading.
ITQ4 The uncertainty would have been half as large, i.e., about 12%.
ITQ5 **(a)** Case (iv); **(b)** (i) 10; (ii) 0.01; (iii) 0.0001; (v) 10; **(c)** greatest in (i); (v); least in (iii).
ITQ6 384 400 km; 384 400 km. First value is more precise; limits of error are ±50 km.
ITQ7 0.5 cm
ITQ8 **(i)** 0.142 mm; **(ii)** 0.142 mm ± 0.0005 mm
ITQ9 **(i)** All resistors marked in this way will have a resistance within the range 9.8 ohms to 10.2 ohms.
(ii) Yes, since the degree of tolerance is an indication of the spread of the values or how different the smallest is from the largest.
(iii) Yes, since it means that the values are all very close to one another.

Examination-style questions

1 (i) A digital clock gives one 'tick' every second. What is the degree of precision of the clock?

(ii) A stop-clock gives 50 'ticks' in 25 seconds. What is the maximum uncertainty in the reading of a time interval measured with this clock, disregarding reaction error? What are its limits of uncertainty?

(iii) A student measured the time for 10 oscillations of a simple pendulum as 14.0 s, using the stop-clock mentioned in (ii) above. The student has a reaction time of 0.2 s and she is known to have started and stopped the watch too late.

(a) What is her maximum reaction error in taking the measurement for 10 oscillations and for 1 oscillation?

(b) What is the maximum precision error in the time for 10 oscillations and for 1 oscillation?

(c) Calculate the total error in the time for 10 oscillations and for 1 oscillation.

(d) Hence write down the time for 1 oscillation, to an appropriate number of significant figures.

2 A sample of a crystalline chemical compound is found to have a mass of 4.2 g when weighed on a certain balance. After being heated for a very long time to remove the water of crystallisation, the compound is found to have a mass of 3.9 g.

(i) What is the precision of the balance used?

(ii) Write down the limits of error of the readings.

(iii) Calculate the mass of water driven off during heating with its limits of error.

(iv) Comment on the suitability, or otherwise, of this balance for an experiment such as this.

3 I measured the inside diameter of a cylinder as 3.42 cm and the outside diameter as 4.60 cm.

(i) Calculate the maximum thickness and the minimum thickness of the cylinder wall.

(ii) State the value of the thickness with its limits of error.

(iii) Write down the thickness to an appropriate number of significant figures.

4 The width of a roughly rectangular field is measured at five different points along its length, and the length is measured at five different points along its width. The values obtained were as follows:

Width/m	24.2	24.2	24.0	24.4	24.0
Length/m	39.8	40.0	40.0	39.9	40.1

Calculate a value, to an appropriate number of significant figures, for:

(i) the mean width;

(ii) the mean length;

(iii) the perimeter;

(iv) the area of the field.

5 (i) Write down the range within which each of the values represented below might be considered to lie:

(a) 642 m

(b) 6000 nm

(c) 3000.0 K

(d) 0.003 J

(e) 1.23×10^{-3} mg

(ii) What is the maximum error (or uncertainty) implied in each of the values given above, as a percentage of the value?

(iii) Which of the values is:

(a) the least precise?

(b) the most precise?

6 John was asked to find the diameter of a very fine wire using a 30 cm rule graduated in millimetres and a piece of glass tubing. To do this, he wound a length of the wire over the glass tubing closely and tightly so that the turns touched one another. He then measured the distance occupied by 42 turns of the wire along the tube as 14 mm.

(i) What was the value of the smallest subdivision of the rule?

(ii) Express the diameter of the wire to an appropriate number of significant figures.

7 (i) Express:

(a) 0.0002 g to the nearest mg

(b) 0.00742 g to the nearest 10^{-2} g

(c) 0.39 mg to the nearest microgram

(d) 4.6 cm^3 to the nearest 0.5 cm^3

(e) 1623 kg to the nearest 5 kg

(f) 16 249 m to the nearest km.

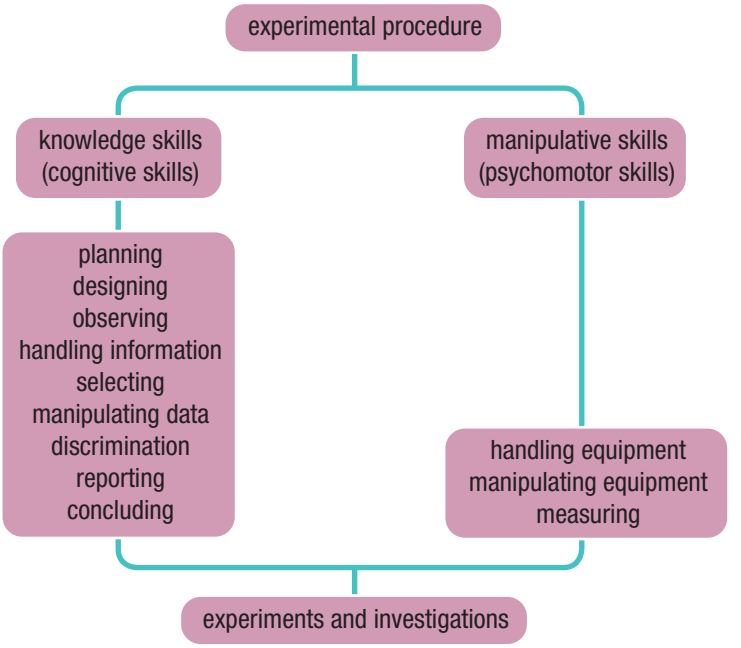
(ii) Suggest a situation to which each of the answers might apply.

(iii) Each of the values above was the reading given by an instrument that measured to the degree of precision indicated. Calculate the maximum percentage error in each of the measurements and hence arrange the measurements in order of decreasing precision.

3 Acquiring Experimental Skills

By the end of this chapter you should be able to:

- understand the nature of parallax errors and systematic errors, and how to reduce their effects
- understand the nature of the skills you will develop by doing practical work
- plan, design, carry out, record readings and compile reports on practical activities
- understand the skills associated with:
  - observation, recording and reporting (O/R/R)
  - measurement and manipulation (M/M)
  - planning and designing (P/D)
  - analysis and interpretation (A/I)



Introduction to experimental skills

In your physics course you will be covering the basic principles of the subject, and you will also be expected to do practical work, some of which will be assessed by your teacher to help determine your final grade.

The practical work that you will do is intended to help you to:

(i) gain the manual skills needed to handle physics equipment in the correct way;

(ii) see the principles and laws of physics put to the test, i.e. to see whether your findings from practical work are supported by theory;

(iii) learn the skills of investigation, such as how to plan an investigation, how to design and execute it and how to use your results to arrive at conclusions regarding your hypothesis.

The marks awarded for your school-based assessment (SBA) projects will be submitted to your examiners for moderation and will form 20% of your overall examination score. You will be awarded marks for skills that you display in performing your practical exercises. Many of these skills will be assessed on the basis of the reports you submit to your teacher. It is important, therefore, that you get as much practice as you can in developing these skills, both in carrying out the practical activities and in reporting on them. In this chapter you will find some useful guidelines and examples of activities that help you develop these skills. You will also find tips that can be used as a guide to doing a self-assessment of your skills.

The question of experimental errors is one that will be present at all times as you carry out your practical activities and, since you cannot always eliminate them, every effort must be made to reduce their effect on your final results.

Experimental errors

Errors committed in the course of doing practical work are of three main kinds:

(a) random errors;

(b) systematic errors;

(c) zero errors.

Random errors

These errors are due to limitations on the part of the experimenter. They are described as 'random' because, in taking a reading on an analogue instrument, an experimenter may as likely as not take a reading that is too high as one that is too low. It is also true that two persons asked to take the reading of an analogue instrument will most likely produce different readings. One reason for this could be an error known as '**parallax error**'. Such an error is caused by the person reading an analogue scale:

parallax error ›

- not using only one eye to view the scale, and
- not looking along a line which is perpendicular to the scale – a requirement if parallax is to be avoided.

Parallax errors and how to avoid them

'Parallax' means 'separation'. When we look at a pointer above a scale and move our eye from side to side, the reading on the scale seems to change (figure 3.1). To make sure that this error is reduced to almost zero, we can use a scale with a strip of mirror in it. If the pointer and its image are in line (as at position B), we must be looking at right angles to the scale, so the parallax error is zero. We can also make sure that the pointer and the scale are very close together. Figure 3.2 shows how this reduces the parallax error.

You may have met parallax errors in chemistry when reading the volume of a liquid in a measuring cylinder. Unless your eye is level with the meniscus, the reading is not quite what it ought to be, as shown in figure 3.3.

Ways of reducing parallax errors

- Use a scale whose pointer moves over a scale provided with a mirror.
- The reading is taken in that position where the image of the pointer in the mirror cannot be seen by the experimenter. Use an indicator that moves very close to the scale (as in a mercury thermometer, a newton-meter and a spring balance). Use the rule on its edge with the calibrations almost touching the object to be measured (as in the case of the measurement of a length).
- If reading the position of a liquid meniscus, ensure that your eye is on the same horizontal level as the lowest point of the meniscus.

CHEMISTRY: measurements

Reducing random errors

Another way to reduce the effect of random errors is to take the average of a number of readings of the quantity being measured. For example, using a tape measure to measure the side of a cardboard box, a pupil obtained the readings in cm shown in table 3.1.

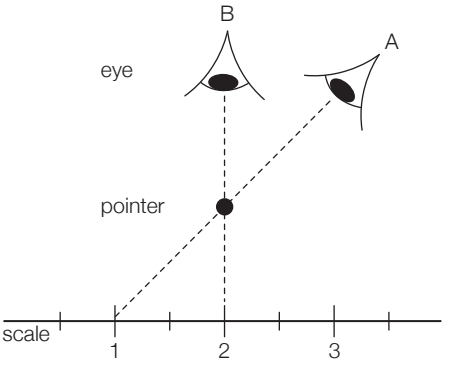


Figure 3.1 Parallax error. The diagram shows the scale of a meter. The dot above it represents the pointer. The eye at position A sees the meter reading as 1, but the eye at position B sees the reading as 2. For the eye at position B, we say that there is no parallax, but for the eye at A there is parallax, i.e. a parallax error.

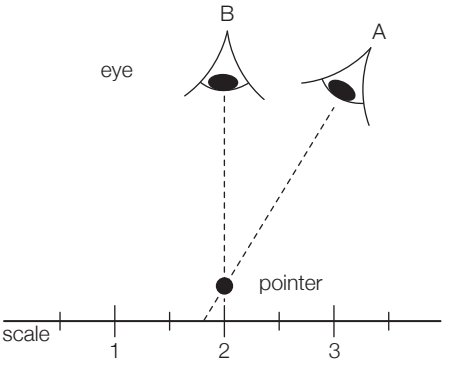


Figure 3.2 The same scale but with the pointer nearer the scale. At the same angle with the vertical the degree of parallax is less than in figure 3.1 Whereas the reading in figure 3.1 was 1 (error is 1 unit), the reading in this case is 1.8, giving an error of 0.2. The correct reading is 2.0, which would be obtained in both cases if the eye was directly above the pointer.

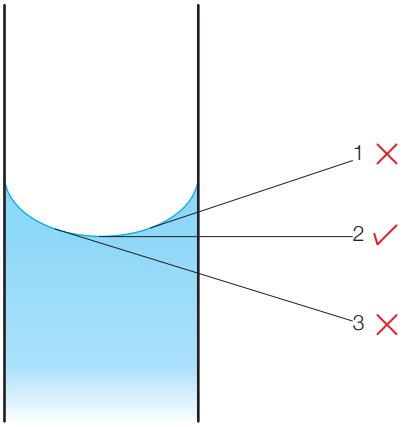


Figure 3.3 To reduce parallax error, take the reading with your eye level with the meniscus, at position 2.

Table 3.1

1	2	3	4	5
63.2	63.4	63.2	63.1	63.2

The average value is 63.22, which is 63.2 cm to 3 significant figures. What does the degree of significance tell you about the calibration of the tape measure that was used?

If the result of an experiment or investigation depends on the relationship between two variables, it is always a good idea to plot a graph between these variables and then to draw the line of best fit among the points obtained from the plot. By plotting a graph of one quantity (normally a dependent variable) against the other (called the independent variable), the effect of random errors may effectively be reduced and a better value obtained for the ratio (called the slope, or gradient) of the dependent variable to the independent variable, than if the value was calculated for each pair of these variables and a mean taken of all the values. Drawing a graph also makes clear the presence of systematic errors of measurement or errors present in measuring instruments.

Drawing a line of best fit among plotted points

Often in order to arrive at a conclusion in an investigation you have to draw a graph of one variable against another. In plotting the points of the graph we would represent each pair of co-ordinates by a dot inside a circle or a cross. Every one of these points plotted is subject to error – error due to imperfect reading of the instrument as well as error possibly due to faulty positioning of the point. Even if we assume that the plotting error was negligibly small, there would still remain random reading errors in the plotted points.

The points plotted will be expected to show a trend which represents the relationship between the plotted dependent and independent variables. These plotted points should not be a jumble of points, but rather a set of points which showed a distinct trend. In ideal circumstances we should find that all the plotted points lie on the same line. We do not, however, conduct experiments in ideal circumstances and since all the points plotted have

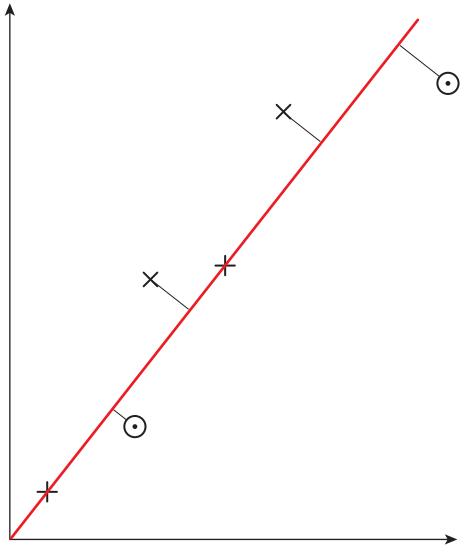


Figure 3.4 An example of drawing the line of best fit on a graph. The line of best fit cancels out the positive and negative errors.

errors, some positive and some negative, some small and some large, it seems reasonable that if we added all the positive errors to all the negative ones, our result should be an error that is pretty near to zero. Thus by drawing such a line where the principle of cancelling out positive and negative errors was followed, we end up with what we call 'the line of best fit' (or 'the best line'). This line should not have too much error. How well drawn it is clearly depends on the degree of care taken in drawing it. It is not necessarily a matter of having the same number of points above and below the line, but rather the total perpendicular distance from points on one side of the line being the same (as far as one can judge) as the total perpendicular distance of points on the other side of the line. In applying this principle, we may well find that the line does not pass through any of the points and, also, that there are not the same number of points on the two sides of it. This is the principle that was used in drawing the line in figure 3.4.

Clearly in drawing such a line, one will have to use a transparent straight edge.

Remember that relationships in physics are not always linear and that we can also have relationships which are represented by curves of many different shapes. The same principle will apply in drawing the best curve among plotted points that show a curved trend.

Finally, it will be appreciated that this method of obtaining the best line becomes more and more valid as the number of points plotted increases. This is why we should always make as many observations (pairs of dependent and independent variables) as is practicable.

Systematic errors

The zero error is the difference between the observed reading (denoted by O) and the true reading (denoted by T).

zero error

Systematic errors are often found in experiments in which analogue instruments are used. They are due, not to the experimenter, but to instruments having a built-in defect. Here is a worked example dealing with **zero errors**.

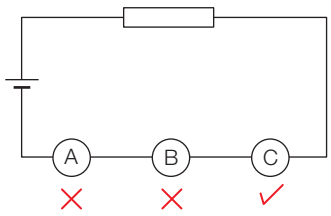


Figure 3.5 The d.c. circuit for Worked example 3.1.

Worked example 3.1

Imagine three ammeters in series in a d.c. circuit, each showing a different current reading (figure 3.5). Meter A reads 3.5 A, meter B reads 2.7 A and meter C reads 3.0 A. Ammeter C is known to read correctly, but A and B are defective. They are said to have zero errors. What is the error in each of the meters A and B?

Solution

We have that

zero error = observed reading – true reading

$E = O - T$

Now, substituting, we have for ammeter A

$E_A = 3.5\,\text{A} - 3.0\,\text{A} = 0.5\,\text{A}$

and for ammeter B

$E_B = 2.7\,\text{A} - 3.0\,\text{A} = -0.3\,\text{A}$

So the answers are as follows:

Zero error in ammeter A = +0.5 A

Zero error in ammeter B = –0.3 A

This means that ammeter A gives a reading that is 0.5 A too high, and ammeter B gives one that is 0.3 A too low.

Zeroing instruments

zeroing instruments ›

Analogue instruments can be 'zeroed' before use. To do this, we need to set the indicator on the instrument to read zero when the value of the quantity being measured is zero. **Zeroing** is necessary when using analogue instruments such as ammeters, voltmeters, newton-meters, lever balances and chemical balances. Each of these instruments is usually provided with a screw device that makes zeroing possible. If the instrument cannot be zeroed, the zero error must be found before the instrument is used. This error, which may be positive or negative, is then subtracted or added (depending on the sign of the error) from the observed reading in order to correct it.

Frequently we do not 'zero' a chemical balance if we are finding the mass of the contents of the vessel, because masses are usually found 'by difference'. A container is first weighed empty, and then weighed with some substance in it. The mass of the substance is found by subtraction, and so the zero error cancels itself out. As seen before, however, it is not always a good idea to use this method of weighing by difference if the mass required is small.

Two instruments in which zeroing is not easily carried out are the vernier caliper and the micrometer screw gauge. In such cases we find the mean of three or four zero errors before the measurement is made, and again after the measurement is made, and then take the mean of the two means as the correction to be applied.

ITQ1

Sketch the scale of the two meters A and B in Worked example 3.1 above. Show on each scale on which side of the zero the pointer is in each of the meters when there is no current flowing in them.

ITQ2

Describe how an error at the lower fixed point of a mercury-in-glass thermometer can be corrected.

ITQ3

The graph showing the relationship between voltage against current for a resistor kept at a constant temperature is a straight line (OX) through the origin, both voltmeter and ammeter having no zero error. Show on the same axes how the graph would change if (i) the voltmeter had a positive zero error, (ii) the ammeter had a negative zero error?

Carrying out practical activities

We consider now the important ideas involved in carrying out practical activities, whether it is for assessment of practical skills or for routine practice.

A short glossary of practical skills

The skills involved in carrying out practical work fall into two main categories:

- cognitive skills;
- psychomotor skills.

cognitive skill ›

A **cognitive skill** (reasoning skill) is one that is based on knowledge or the use of knowledge. An example of this skill is that needed to decide what should be the range of an ammeter to be used in an electricity experiment.

psychomotor skill ›

A **psychomotor skill** (manual skill) is a skill whose application requires the use of muscles of one or more kinds. In physics these are generally muscles of the hand and the eye. An example of this skill is that needed to set up a retort stand (involving the use of the hand) and to locate an optical image formed by a lens (involving the use of the muscles of the hand and also the eye).

The following definitions should help you to understand the meaning of each skill mentioned in the SBA part of your syllabus. The skills are divided into three categories. The SBA code for each is shown in brackets, i.e. (X/X).

- *Knowledge and comprehension* (K/C). This means remembering things and using that remembered knowledge in familiar situations.
- *Use of knowledge* (U/K). This means using facts and procedures in unfamiliar situations; handling data; making judgments; making predictions; putting ideas together; recognising limitations; and making reasoned judgements.
- *Experimental skills* (X/S). These skills are needed in order to follow instructions in doing practical work and in deciding what observations to take and how to record them; prepare tables, draw graphs and diagrams and present reports in an organised manner.

Also included in the X/S group is the ability to design experiments by yourself for a particular purpose and, in doing so, to realise limitations and dangers and to modify your design, if necessary, in the light of experience. Another skill that might be included in this category is the ability to analyse results of measurement and to decide on matters such as limits of error and significance.

We shall now look more closely at the skills in the X/S group. In doing so we shall be considering a number of activities that will lend themselves to an examination of the experimental skills identified in your syllabus, such as observing, analysing, interpreting, measuring, manipulating, recording and reporting. First, however, we briefly explain what each of these skill terms denotes.

Analysing, interpreting and handling experimental data

This means using your knowledge and understanding of physics theory and your mathematical skills to help you arrive at a conclusion. This may require you to:

1 manipulate equations;
2 draw a graph;
3 extract information from the graph.

Designing an activity

This means deciding, in some detail, how an experiment or investigation can be conducted in order to achieve a stated aim. A design would contain:

1 the list of apparatus to be used;
2 how the apparatus is to be used, what readings are to be taken and how they are to be used;
3 how a conclusion would be reached;
4 an assessment of how reliable the conclusion is.

Handling apparatus

This skill is concerned with the way in which apparatus is used in an experiment. The results of an experiment or investigation are most reliable when the readings obtained from the observations made show a high degree of accuracy and precision. In doing experiments, therefore, we should always use apparatus in such a way as to achieve both of these. Accuracy and precision have been discussed in chapter 2.

CHAPTER 2

Manipulating apparatus

This is slightly different from handling apparatus, in that it might involve, for instance, adjustment of the position (or the 'setting') of a piece of apparatus in order to produce a particular result.

Observing and observation

An observation could also mean the set of readings you obtained from a certain setting of your apparatus through the use of one or more of your senses.

This is using one or more of the five senses (sight, hearing, smell, touch and (rarely) taste), to obtain a reading (make a measurement) or to detect a difference of condition of some kind. Examples are looking at an image, reading a thermometer and timing oscillations.

Planning

This means considering and deciding what should be done in preparing an experiment with a view to achieving a stated aim. A plan is usually a broad

outline of what is to be done. Its structure is generally based on the theory related to the aim of the activity. A plan is, essentially, what you are going to do in order to arrive at a conclusion to your investigation.

Recording

This is putting down the observed value of a quantity with its units and with an appropriate degree of significance, for example, on a table, or in a line graph.

Reporting

This means writing a full account of all the equipment used and the steps carried out in performing a practical activity, together with all the stages used in arriving at a conclusion, as well as any condition that existed in the course of the activity which might have had an adverse effect on the outcome of the activity.

Experimentation

Experiment or investigation?

experiment ›

It is very likely that most of the practical work you will do will involve finding the value of a physical quantity. We call this an **experiment**. However, there are going to be cases where you will be asked to 'Find out whether …' or to 'Find out how …'. For example, you may be asked to find out whether the extension of a rubber band is proportional to the force stretching it. If the activity is to find out 'whether …' or 'how ...', we call it an **investigation**.

investigation ›

aim conclusion ›

Whether the activity is an experiment or an investigation, however, there will always be an **aim** and a **conclusion**. The aim is an opening statement of the purpose of the activity. The conclusion is a closing statement of the result or the outcome of carrying out the activity.

An activity might be based on a familiar situation such as finding the period of a pendulum of given length or the sizes of the feet of your classmates) or a situation that is less familiar and whose results are harder to predict. For example: What is the effect on the period of a pendulum if we keep themass of the bob constant but alter its radius? Is there a relation between the height of males and the size of their feet?

For any activity, experiment or investigation, the conclusion should always be related to the aim.

Undertaking a practical activity may be divided up into three distinct parts:

1 Preparing, i.e. planning and designing the activity.
2 Executing, i.e. carrying out the activity.
3 Reporting, i.e. compiling an accurate account of all that was done in relation to the activity and stating the result.

Preparing

preparing ›

To prepare an experiment or an investigation means to plan it and then to design it. To plan it is to **prepare** an outline of what you are going to do, whereas to design it is to draw up a detailed account of how you are going to carry out the plan. The plan will be based on the theory associated with the activity. The **design** will go much further. It will be concerned with all or most of the following points

design ›

1 Deciding on the list of apparatus to be used, and a diagram showing how the apparatus is to be arranged.

range ›

2 The **range** of values to be employed, if appropriate.

sensitivity ›

3 The range and **sensitivity** of the items of apparatus (see figure 3.6).

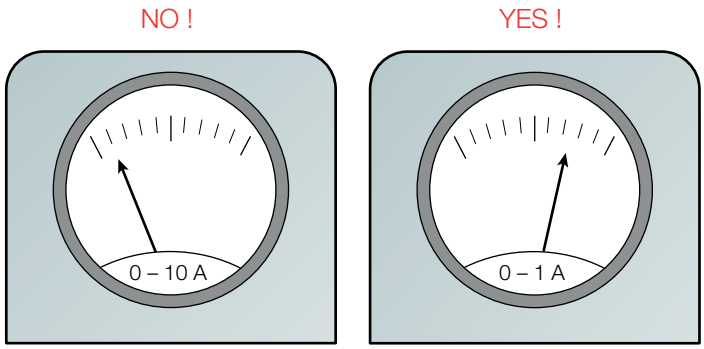


Figure 3.6 An example of a decision you might have to make when planning an experiment. Why is the 1 A meter a better choice?

The sensitivity of an instrument is the value of the quantity measured by that instrument, which is represented by 1 scale division. For example, if there are 100 whole scale divisions on the scale of a newton-meter that measures forces up to 10 N, then 1 scale division would represent $\frac{10}{100}$ N or 0.1 N. The sensitivity of this newton-meter would therefore be 0.1 newton per division or 0.1 N div^{-1}. Sensitivity is not the same thing as precision or accuracy (see chapter 2).

4 Precautions to be taken to achieve the best accuracy and to ensure against accidents and damage to any person or to property.

5 The sizes of quantities that might make for precision in your readings, where possible.

limitations ›

6 **Limitations** are circumstances or conditions that might prevail during the conduct of the activity that might adversely affect readings obtained. Very often these conditions are beyond the control of the experimenter, and steps must be taken to reduce their effects.

Try to get into the habit of presenting your results in a neat and orderly manner. Do not scatter results all over the page.

7 The presentation of your results. This is an important consideration. Should the results be tabulated? If so, how should the table be drawn up and what should the headings be? What should be the degree of significance in each column?

8 The question of whether or not a graph should be drawn and, if so, of what variable against what other variable. If a graph is drawn, some use will be made of it. In many cases a slope is taken. If this is the case, the design will include the way in which this slope is to be used to help you to arrive at a conclusion. The steps between the graph and its use in arriving at a conclusion will all have to be carefully considered in the design.

variables ›

The **variables** are the quantities that change as the experiment or investigation proceeds. One quantity is given a convenient value to start with. (This is called the **independent** variable) and as a result another quantity in the activity changes. This quantity is called the **dependent** variable. When we draw graphs, we usually plot the dependent variable on the vertical axis (see figure 3.16).

independent variable ›
dependent variable ›

Executing

Two helpful hints

1 Analogue meters must be chosen so that maximum readings obtained on them occupy a fairly large fraction of the full range. If a digital meter is used, the range selected should give as large as possible a number of significant figures in the reading (better precision).

2 Never take your apparatus apart before you are satisfied that your work was brought to a satisfactory conclusion. You may need to use it for a further check!

After you have drawn up your plan and your design, you are ready to carry out the activity. This should be done in two parts:

- the first part should be a trial (sometimes called a 'dry run');
- the second part is the activity proper.

To carry out the dry run, the apparatus is set up in accordance with your design. If, during this dry run, you find that it is necessary to make changes to your design, then you should feel free to do so. You must be sure, however, that the changes you introduce would produce a better result at the end.

If the dry run goes as expected and you encounter no hitches along the way (such as conditions you did not anticipate and therefore made no allowance for), then you will be ready to carry out the 'real' experiment.

In doing so, you will now follow the same steps as before, taking more care this time to measure quantities more carefully. If the activity is one in

Golden rules

1 When you make measurements, never record the readings on odd bits of paper, or on the palm of your hand, or in an exercise book that is used for a subject other than laboratory work in physics. All readings taken must be recorded in your practical notebook.

2 Do not try to commit readings to memory.

3 It is better to record readings in ink. Ink does not fade like pencil.

smooth curve ›
A smooth curve is a graph without kinks.

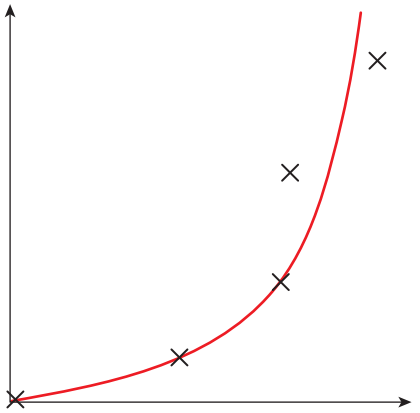


Figure 3.7 Draw a smooth curve through the points.

which a graph is to be plotted, this graph should be plotted as the values of measurements are obtained. Axes for the graph can be set up once the maximum and minimum values of the independent variable are used in the trial. Be sure to check that the corresponding values of the dependent variable can be measured on the instruments selected for use in the experiment or investigation. If this is not the case, you will have to adjust the range of values you use to suit your instruments or to change the range of the instruments to suit the range of values envisaged in your plan. Plot these two extreme points on the axes.

Having plotted these two points, you should now plot the intermediate values and look for the shape of the emerging graph. You should find that your graph is either a straight line or a **smooth curve** (figures 3.7 and 3.8). You may find that a point or two do not quite 'fit in' with the trend that the graph shows. In a case like this, you should go back to your apparatus and check the co-ordinates of the point again. If you get the same values, then you should go ahead and plot them. Do not reject the point at this stage.

Sometimes plotted points are so arranged that it is difficult to tell whether the graph to be drawn among them should be a smooth curve or a straight line. Guidelines that will help you to decide how to draw the best line or the line of best fit are given on pages 28 and 29).

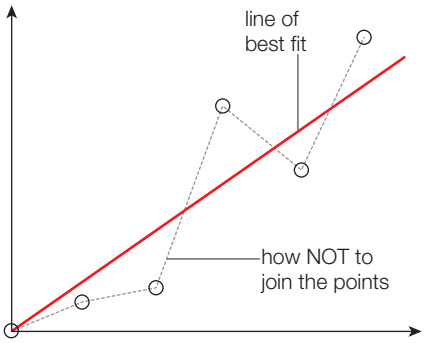


Figure 3.8 Draw the line (curved or straight) of best fit. Do not join the points.

Reporting

report ›
Your **report** should be an accurate, dated account of all that was done in carrying out your practical activity. If the activity is an experiment, you can use some or all of the following headings.

The aim

This is a statement of the purpose of the experiment. The method to be used should be stated, where possible (for example: 'To determine the density of a liquid by Archimedes' principle').

Apparatus

You should write down all the items of apparatus used in the experiment, together (where possible) with a statement of particulars of each (e.g. range and precision for ammeters, thermometers and newton-meters).

Labelled diagram

Perspective (three-dimensional) diagrams do not add value to your report. They take longer to draw and look more complicated.

Draw a labelled diagram of the experimental arrangement that was used. This diagram would normally be a large line diagram, drawn as a vertical section. There should be no shading.

Method or procedure

This is a full, but brief and dated, account (in prose form) of all that you did using the equipment. You should list your steps in the correct order. There should be no results of measurement in this account. Precautions taken to

improve accuracy and to ensure safety to any person and to equipment must be included.

Results

These should be given in a table if the value of an independent variable was varied in the course of the experiment. The columns or rows used for recording the values should be appropriately headed with the name of the quantity concerned, a capital letter to denote it and the unit being employed. In expressing the unit, solidus notation should be used (e.g. current, I/A). Each entry should have an appropriate number of significant figures, and each one relating to a given quantity should show the same degree of significance. No mathematical operations should be done in the table.

Example of theory

For a simple pendulum:

period $T = 2\pi\sqrt{\frac{l}{g}}$

$T^2 = (\frac{4\pi^2}{g})l$

In a graph of T^2 against l,

slope = $2\pi/g$, a constant

Theory

The theory upon which the experiment is based should now be given. If your plan and design included a graph, this statement of the theory would lead up to the equation of the graph. If the slope of the graph or either of the intercepts was to be used in arriving at a conclusion, this would be explained under this section.

Graph

When deciding on suitable scales, avoid using the numbers 3, 7, 9, 11 as they are awkward to use as divisors. 2, 4, 5, 10, 20 are much better numbers to use.

The scales chosen for the axes of the graph should be such as to produce a graph that occupies no less than three-quarters of the graph page. The axes should be labelled with the name of the quantity being plotted (or a symbol for this quantity) and the units given, again using solidus notation. Points should be represented either by a dot inside a circle or by a cross. If a dot is used it should be very small. The line you draw should be very thin. For this reason, a sharp, hard pencil should be used to plot the points and to draw the curve. This curve should be smooth, as mentioned earlier. We never join points by straight lines in physics (see figures 3.7 and 3.8).

Calculations

inference ›

The theory given above would have led to the result to be used in arriving at a conclusion. This result is now applied to the graph and a value extracted or an inference made from the shape and behaviour of it. This result (or **inference**) will be of value in arriving at a conclusion.

Discussion

A possible limitation in a pendulum experiment might be a constant draught in the laboratory. You might reduce the effect of this draught by:

- using very small swings;
- carrying out the experiment in a far corner of the laboratory.

A suggestion for improvement might be to suspend the pendulum from the ceiling rather than from a clamp in a retort stand on a laboratory bench.

What is (or are) the advantages of using a very long pendulum?

Here mention is made of limitations encountered in carrying out the activity and what steps were taken to minimise their effects. You might have thought, in carrying out the activity, that it might have been better to adopt a different approach in a certain area. For example, you might have thought that the experiment could have been improved with a slightly different feature of the design from the one you used. This would be a good place to mention this. It would be a good idea, therefore, to have two subheadings under this general heading, namely:

- limitations;
- suggestions for improvement.

Conclusion

Remember the following:

- The conclusion you arrive at must be related to the aim you set yourself.
- The report you write must be your own work even where you have joined with others in carrying out the activity.

Here you use the evidence collected from the measurements made and from the graph (if applicable)

either (for an investigation)

to state whether the aim you stated at the beginning was achieved, and, if it was, what was the outcome,

or (for an experiment)

to calculate the value of the quantity in question.

Examples of possible experiments for testing practical skills

Having now explained the procedure to be followed in doing an experiment or an investigation, we now begin our discussion of possible practical activities which you can carry out to practise or to test experimental skills. We will take four activities, one for each of the four categories of skills identified by CSEC. We take first an experiment based on the extension of a spiral spring, which is to be assessed for skills in observing, recording and reporting (O/R/R).

Practical activity 3.1

The aim, or purpose, of the activity is clearly stated.

Investigating the extension of a spiral spring under various loads

In this activity we will examine the relationship between the load applied to a spring (or the tension in the spring) and the extension produced in the spring as a result.

You will need:

- retort stand and clamp
- spiral spring
- half-metre rule
- *either* a set of slotted masses with carrier
  or a light, but stiff, cardboard scale-pan with a set of masses of values up to 0.5 kg in steps of 50 g.

Note that the carrier or scale-pan must be sufficiently near to the rule to minimise parallax reading errors, but it should not touch the rule.

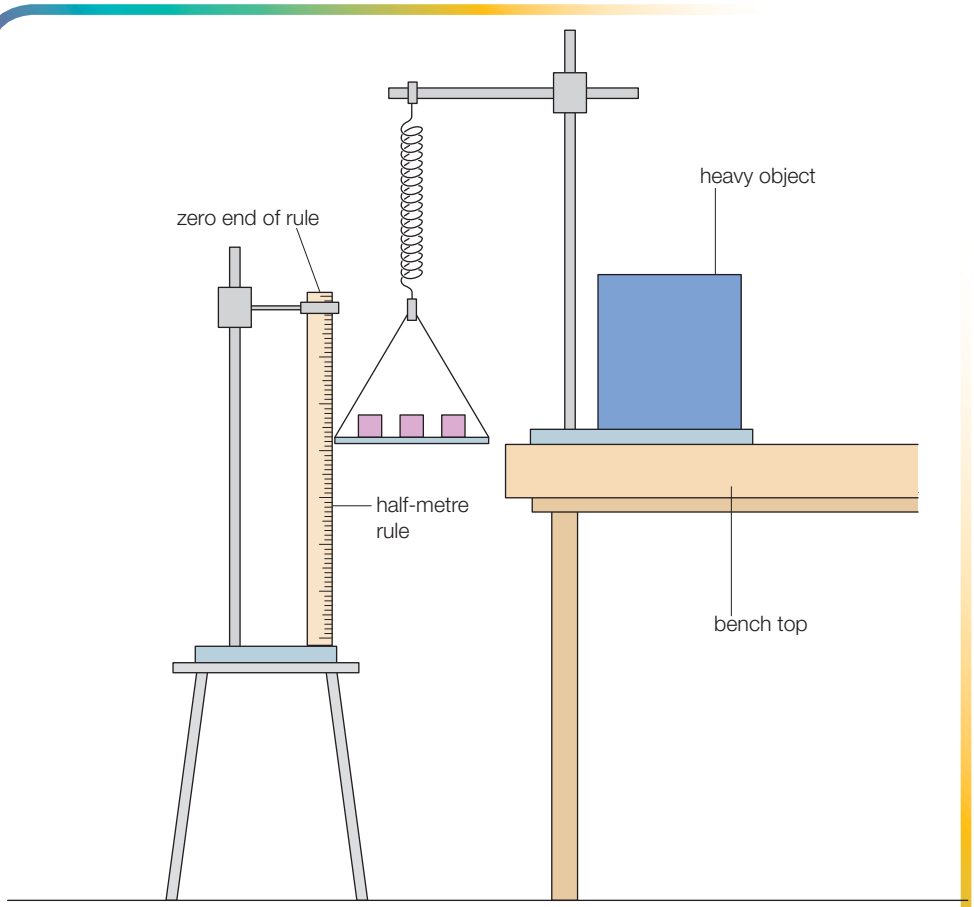


Figure 3.9 Apparatus used to investigate the stretching of a spring under load.

Method

1 Set up the apparatus as shown in figure 3.9, but do not hang the spring on the clamp just yet.

2 Measure the original (or natural) length of the spring by laying it on the half-metre rule (lying on the laboratory bench) and measuring the length with the spring stretched out straight, but not pulled to produce tension. This length will be the distance from the start of the first coil to the end of the last. Let this length be l_0 (mm). Record its value.

3 Now securely attach one end of the spring to the clamp of the stand so that it hangs vertically and clear of the edge of the laboratory bench (see figure 3.9) Attach the mass carrier to the lower end of the spring. (This counts as the first 50 g mass hung on to the spring.) Read the position of the lower surface of the carrier along the rule. Denote this reading of the stretched length by l_1. Record its value on a table drawn up as shown in table 3.2. If a cardboard scale-pan is used, its mass can be ignored.

4 Set up the metre rule in the retort stand as shown in the diagram ensuring that it is vertical and is just not touching the scale-pan.

5 Do a 'dry run' by loading the spring up the maximum of 500 g, then unload it in steps of 50 g back to zero.

6 At this load of zero, check to see whether the unloaded spring still has the same unstretched length, l_0, as before. If it does, then the actual experiment can now be started. If it does not, but the length is greater that it was before, it would mean that the spring has been overstretched because it was overloaded. Another dry run will then be needed until it is clear that the spring will not be overloaded by any of the loads to be used.

7 Now start loading the spring once more from 50 g by adding masses to the carrier in steps of 50 g to increase the load or the tension in the spring. Each time a mass is added, note the position of the carrier along the rule. Record the values of load and stretched length, l_1, in a table like the one shown in table 3.2. Increase the load up to the maximum value reached in the dry run.

Table 3.2 Unstretched length of spring, l_0 = … mm

Mass used, *m*/g	Load, *F*/N	Stretched length of spring, l_1/mm			
		Load increasing	Load decreasing	Mean stretched length, l_1/mm	Mean extension $(l_1 - l_0)$/mm
50	**0.500**				
100	1.000				
150	1.500				
200	2.000				
300	3.000				

Plotting the graph

Now plot the graph of the load, *F*/N against the mean extension, $(l_1 - l_0)$/mm, starting both axes from 0. If your graph is a straight line through the origin then it shows that the load applied (or the tension in the spring) is proportional to the extension produced in the spring by that load.

Conclusion

Assuming that the graph is a straight line through the origin, your conclusion should be that the extension of the spring is proportional to the load *or* that the load applied to the spring is proportional to the extension.

Discussion

Your discussion might be based on the answers to the following questions:

(i) Were there any conditions that arose which made it difficult to proceed or which gave rise to reading error?

(ii) Could the two sets of extension obtained have been used in a different way? If so, what is that way? Which method of treatment would be better? Why?

(iii) What information could be extracted from the graph obtained?

The answers might be:

(i) Provided that precautions were taken as mentioned in the method, there should have been few adverse conditions to cope with. Parallax errors are the most likely hitches.

(ii) Two sets of points could have been plotted, namely:

(a) the set obtained by plotting load and the extensions under increasing load, and

(b) the set obtained by plotting load and the extensions under decreasing load.

There would in this case have been twice as many points plotted. Do you think that this would have been an advantage and would have led to a better straight line of best fit than the one obtained by using the mean extension? Would having to draw the best line through twice as many points have been an advantage or more difficult to do? Give a reason for your answer.

(iii) The quantity known as 'the spring constant' can be obtained from the graph. This constant is defined as the force or tension needed to produce unit extension (i.e. 1 cm or 1 m) in the spring.

As long as a helical (or spiral) spring is not overstretched, it will have a spring constant. Stiff springs (such as those used in car suspensions) will have constants of very high value. Soft springs (like those used in ball-point pens) will have constants of very small value.

The graph you obtain in the experiment described will most likely be a straight line through the origin. Suppose, in a similar experiment in which a dry run was not done, a graph such as that in figure 3.10 was obtained.

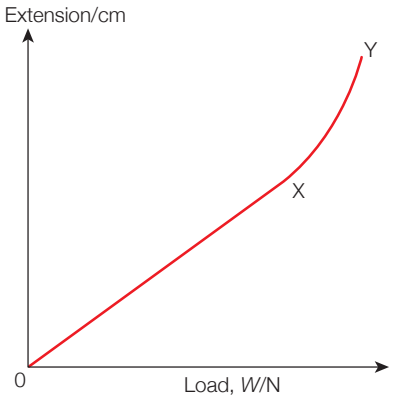


Figure 3.10

(i) What would be the most likely reason for this different shape?
(ii) Explain why the non-linear part of the curve:
- is curved, and
- is shaped as it is and not curved the other way.

(iii) Make a statement explaining what is happening to the spring in the non-linear portion of the curve.

Criteria for the assessment of practical skills

Assessment criteria for O/R/R skills

For an assessment of observational skills displayed the assessor might ask some or all of the following questions:

Did the student take care to ensure that the conditions required for the success of the activity were met? More particularly, for observational skills, did he/she:

- ensure that the stand supporting the spring could not move in the course of the experiment?
- ensure that the metre rule was always vertical when measurements of the extension were being made? If so how was this done?
- ensure that the edge of the scale pan was as near as possible to the metre scale in order to reduce parallax error to a minimum?
- use a reasonably wide range of values of the load?

The points listed above are all concerned with the observation of events and correct conditions and this would clearly have involved the use of one or more of the senses.

Recording skills

These skills are related to the manner in which you display the readings which result from measurement. Some of the important questions here are the following:

- Were all the readings recorded on a table?
- Were the columns of the table properly headed with the quantity measured, the symbol for the quantity, and with solidus and unit used?
- Were all the readings in a given column expressed with the same degree of significance?
- Was the significance of the readings in any one column in accordance with the precision of the device used to obtain them?
- Was the origin (0, 0) shown on the graph?
- Were the scales chosen for the two variables and the range of the variables plotted such that the graph obtained occupied at least half of the page area?

Reporting skills

Some of the skills expected here are related to the following questions:

- Was the student's report:
  - (a) titled appropriately?
  - (b) dated?
  - (c) in the past tense and in prose form and not given as a set of instructions?
    - Were the grammar and the syntax used of a satisfactory standard?
    - Was there a sectional line diagram showing how the items of apparatus listed were arranged for use? Was the diagram labelled, or correct symbols used where expected?
- Was the sequence of the steps used in the report the correct one?
- Was there mention of difficulties encountered and were there any adverse conditions present which were hard or impossible to control? If there were, did the student state what steps he/she took to reduce the effect of these adverse conditions or these difficulties?
- Did the student show clearly how the readings he/she recorded and the graph he/she drew were used in order to arrive at a conclusion?
- Was the origin (0, 0) shown on the graph where it was necessary?
- Were the scales chosen for the two variables and the range of the variables plotted such that the graph obtained occupied at least half of the page area?
- Was the final result declared related to the aim stated at the beginning?

Further activities that are suitable for the assessment of O/R/R skills can be found in **CHAPTER 37**.



We go now to the practice and assessment of measurement and manipulation (M/M) skills.

Practical activity to practice and to assess manipulation and measurement (M/M) skills

There are a number of activities which will lend themselves to the assessment of these skills, since almost all physics experiments involve measurement of physical quantities and manipulation of equipment. These skills depend on

manual dexterity, good judgment and good hand and eye co-ordination. This experiment provides an opportunity to use these skills.

Practical activity 3.2

Verifying the laws of reflection

In this activity we will verify the three laws of reflection.

You will need:

- large plane mirror
- two large nails about 6 cm long, or two long pencils
- metre rule
- two retort stands with clamps
- large drawing board
- large protractor.

Method

1 Place the plane mirror flat on the laboratory bench with its reflecting surface uppermost.

2 Place a nail (the 'object' nail) with its head in the clamp of one of the stands and incline it over the plane mirror at an angle of about 45° to the mirror and with its lower tip about 3 cm away from it (see figure 3.11).

3 Look at the image of the nail in the mirror and adjust your direction of view until you are looking along the image and can see only its tip.

4 Place the head of the other nail (the 'sighting' nail) in a clamp and adjust its position and orientation so that, looked at with one eye, it is in line with the image of the first nail (see figure 3.11). When this is done, you should not see the image of the first nail at all as you look along the length of the sighting nail, since the second nail will be covering it.

5 Now carefully place the drawing board in such a position that both nails, still supported in the clamps, touch the board along their entire length.

6 Check (with your protractor) whether the plane of the board is perpendicular to the plane of the mirror.

7 Hold a length of string along the object nail so that one end of the string touches the mirror. Using a large protractor, measure the angle, α, which the string makes with the mirror surface (see figure 3.11). You may need another pair of hands to help with this! Do the same with the sighting nail to measure angle β.

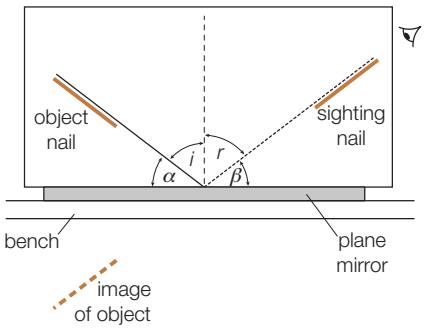


Figure 3.11 Testing the laws of reflection. Side view of arrangement.

Results

If the sighting nail was positioned so that it blocked off the image of the first nail, then it was in line with the reflected ray. If both nails, representing the incident and the reflected rays, were in the plane of the board and this plane was perpendicular to the mirror, then:

- the normal to the mirror was in this plane, and
- the incident ray and the reflected ray were both in the same plane as the normal. This is the first law.

From figure 3.12 it can be seen that if the value of angle α is the same as that of angle β, then the value of the angle of incidence, i, (which is the complement of α) is the same as that of the angle of reflection, r (which is the complement of β). We can determine whether the values of α and β could be deemed to be equal by considering the limits of uncertainty of each angle. If the angles are found to be equal, then we must conclude that the angle of incidence is equal to the angle of reflection and that the second law is also true.

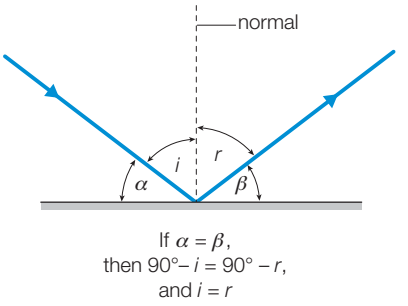


Figure 3.12 The angles α and β are called the 'glancing' angles.

Assessment criteria for M/M skills

In order to assess the student for the relevant M/M skills, an assessor might ask the following questions:

For the assessment of manipulation

- Did the student look with only one eye whenever he/she was checking for alignment of sighting pin and image?
- Did the student (if he/she worked unaided):
  - (a) show dexterity in aligning the image pin?
  - (b) show dexterity in positioning the drawing board?
  - (c) show deftness in measuring the angles α and β without assistance from another student?

For the assessment of measurement

- How did the student decide whether, in the light of the fact that both angles were subject to uncertainty, there was a reasonable chance they could be equal?

Further activities that are suitable for the assessment of M/M skills are given in **CHAPTER 37**.

Activity to assess planning and designing (P/D) skills

Before we begin to plan an activity of this kind, it is sometimes a good idea to do some thinking around the key topic, as shown in the mind map in figure 3.13. This map concerns the activity that will be described below – to determine which material, plastic or newsprint, is the better insulator.

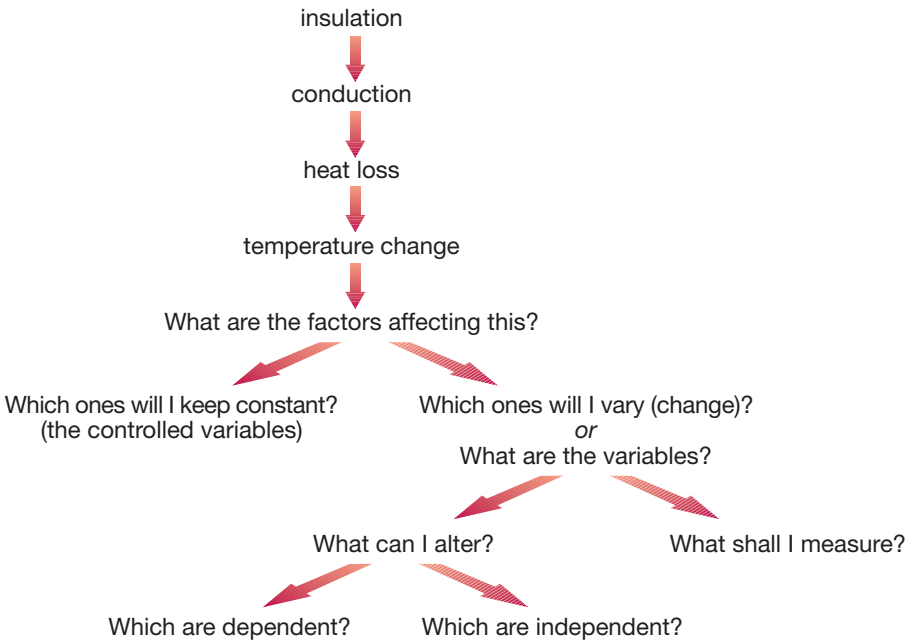


Figure 3.13 A 'mind map' for tackling an investigation.

We are now ready to draw up the plan and the design for the investigation mentioned above.

To investigate whether plastic sheeting is a better heat insulator than newsprint

The topic: the insulating properties of plastic and newsprint

The plan

The plan will be to allow two identical liquids to cool through the same range of temperatures under identical conditions, except that one will be surrounded by plastic sheeting and the other by newsprint. The thickness of the plastic insulation and that of the newsprint must be the same. The insulation surrounding the system that cools more slowly will be deemed the better insulator.

The design

Aim

State the aim. In this investigation, it is the following question:

> 'Is newsprint a better heat insulator than plastic?'

Hypothesis or no hypothesis?

hypothesis ›

Because this investigation is a YES/NO question with only one possible answer, you can state a **hypothesis**. Your reason may be based on your everyday experience or on theoretical reasoning. It may go like this:

> 'I think that plastic is the better insulator, since it is much more widely used as table-mats than materials similar to paper, like straw.'

Don't state such a hypothesis if there is more than one possible answer; in science we must keep an open mind.

Apparatus

The apparatus needed for this investigation might be as follows:

- copper calorimeter with Styrofoam lid, with a hole for a thermometer;
- thermometer, –10 to 110°C (× 0.5°C);
- Bunsen burner with tripod and gauze;
- plastic sheeting;
- sheets of newsprint;
- rubber bands;
- micrometer screw gauge.

Theory of cooling

What are the factors that affect the cooling rate of a hot body?

At this point, we consider what steps we should take to ensure that we are investigating only the effect of the insulation on the cooling rate of the hot liquids. We recall that the cooling rate depends on the following factors:

1 the area of the surface of the copper calorimeter that will be exposed to cooling by convection;
2 whether there are any draughts present in the area of the room where the cooling is to take place;
3 the temperature range over which cooling will take place;
4 the thickness of the insulation used;
5 the amount of water used.

controlling variables ›

If we are to test only the effect of the insulation, we must ensure that the other factors just mentioned are all the same for both systems. This is called '**controlling variables**'. So to 'control variables' for the factors listed above (in the same order):

1 we use the same physical arrangement for each;
2 we ensure that the cooling of the two systems is carried out in the same position in the laboratory;
3 we use the same cooling range for the two systems;
4 we check that the thicknesses of the two materials are the same, and ensure that the plastic sheets and the newsprint sheets are pressed firmly together, with no air spaces between them;
5 we make a mark inside the calorimeter about three-quarters of the way up and fill to this mark each time.

Diagram

Draw a diagram of the experimental arrangement (see figure 3.14).

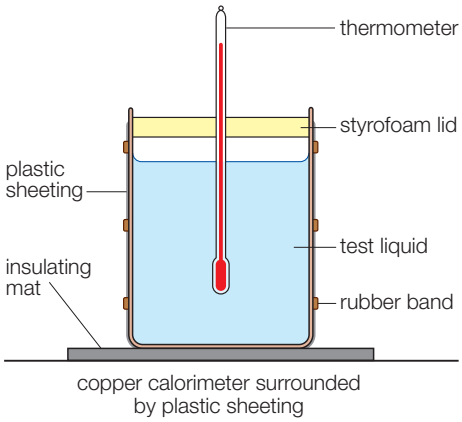


Figure 3.14 Diagram of the apparatus for the investigation.

Method

We are now ready to consider the procedure. The steps are as follows:

1 Place a 500 ml beaker of water over a lighted Bunsen burner. With a micrometer screw gauge, determine how many sheets of newsprint have the same thickness as one sheet of plastic.

2 Tightly wrap this number of sheets of the newsprint round the calorimeter and hold the sheets in place with rubber bands (see figure 3.14). The insulation is used over the entire surface of the calorimeter, except the top, where there will be a lid of Styrofoam.

Take great care with the boiling water!

3 When the water boils, fill the calorimeter with boiling water up to the mark made previously. Pass the thermometer through the Styrofoam lid and cover the calorimeter with the Styrofoam lid, making sure that the thermometer bulb is in the water.

4 Allow the temperature of the thermometer to rise to the temperature of the water then fall slightly. Shortly after the temperature begins to fall begin to record the temperatures of the cooling liquid, shown by the thermometer, at intervals of 1 minute. In a table, record the fall of temperature for about 30 minutes.

5 Repeat steps 2 to 4 using plastic sheeting of the same thickness as the newsprint instead of the newsprint.

Results

Results should be given in a table for each material, as shown in table 3.3.

Table 3.3 An example of how results should be presented.

Newsprint		Plastic	
Time/ min	Temp./°C	Time/ min	Temp./°C
1			
2			
3			
etc.			

How will results be used?

You should plot a cooling curve for each of the two cases. These curves will appear as shown in figure 3.15.

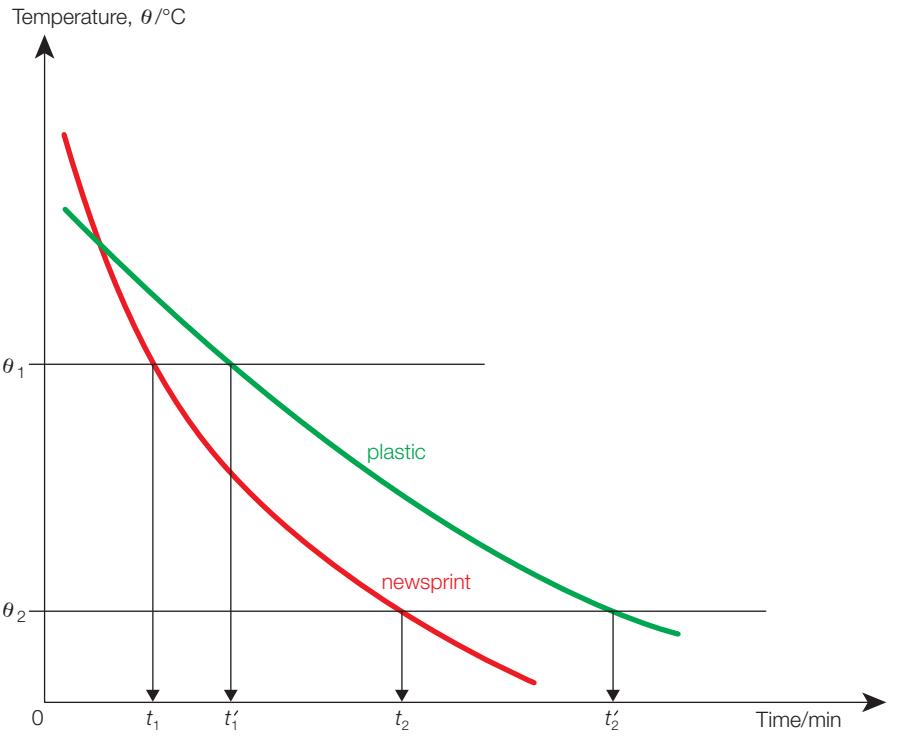


Figure 3.15 The cooling curves from the investigation.

Although the points obtained during the cooling of the two systems may not seem to lie on a smooth curve, a smooth curve should, nevertheless, be drawn among the points. Remember that we do not 'join the dots' in physics, but 'follow the trend'.

By convention, we plot the independent variable horizontally and the dependent variable vertically (figure 3.16).

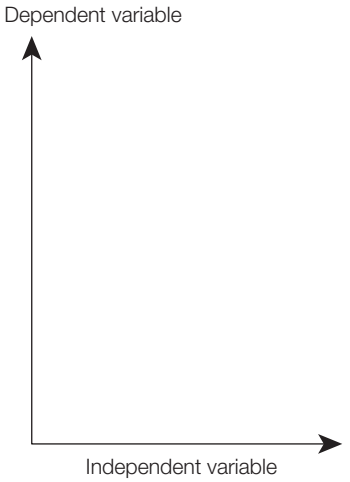


Figure 3.16 Plotting independent and dependent variables.

How will cooling curves be used?

The better insulator will produce slower cooling, and so its cooling curve will fall more slowly than that for the poorer insulator. This can be seen from the graph. Whereas the newsprint curve takes only $(t_2 - t_1)$ minutes to cool through the temperature range $(\theta_1 - \theta_2)$°C, the plastic curve takes the somewhat longer time of $(t_2' - t_1')$ minutes.

Discussion

One point of discussion could be whether we could be sure that the effect of the draught around the two liquids during cooling would be the same, which was assumed although the cooling took place over different periods. Might it have been better to have used two identical calorimeters and observed the cooling of both liquids over the same period thereby ensuring identical draught conditions?

How would a conclusion be arrived at?

If the investigation were carried out and the cooling curves did look as indicated in figure 3.15, then the hypothesis stated at the beginning would be correct. If, however, the graph for plastic was steeper than that for newsprint, then the opposite would be true and the hypothesis would be false.

Your conclusion will therefore depend on what the two graphs reveal.

Assessment criteria for P/D skills

Further suggestions of activities that could be carried out to assess planning and designing skills are given in **CHAPTER 37**.

The points that may be raised in assessing these skills are left for discussion by students and teacher. Some of these points are related to:

- the control of variables;

- the use of a hypothesis;
- the features of the design;
- limitations.

Issues involved in assessing analysis and interpretation (A/I) skills

The skills of analysing and interpreting are required in almost all practical activities we undertake. For example, if we were to carry out a pendulum experiment and were asked to determine how the period of the pendulum was related to the length of the pendulum, and had available a graph of the $(\text{period})^2$ against the length, we would first have to use reasoning skills to analyse the graph to discover what it was telling us about the period and the length. We would then have to use our mathematical skills to infer that the square of the period was proportional to the length (assuming that the graph showed no intercept and was straight). What this graph reveals does not quite answer the question as to how the period was related to the length. All we could say, judging by the graph, is that an increase in the length resulted in an increase in the period, but how much increase there would be for an increase in length by a certain factor would require us to delve into further mathematics of variation. It is only if we know how to interpret direct variation that we would be able to predict that if the length changed by a factor 'f', the period would change by a factor of '$f^{1/2}$'). It would be clear, therefore, that the skills needed for analysis and interpretation are certainly more cognitive that psychomotor, certainly more on the reasoning side than on the manipulative side.

Further activities that require the use of analysis and interpretation skills will be found in **CHAPTER 37**.

A useful exercise from the point of view of A/I skills and P/D skills would be to investigate how the period of a pendulum depends on the mass of the bob and a more challenging adventure would be to find out how the period depended on the radius of the bob. There would be a fair number of variables to be controlled in such exercises.

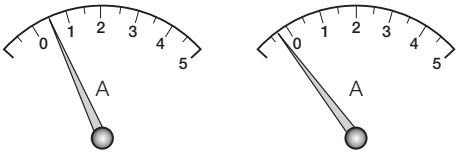
Chapter summary

- A practical activity in physics requires the use of skills of two types: knowledge skills and manipulative skills.
- Knowledge skills are those based on the knowledge of theory and the application of this information.
- Manipulative skills are those based on precise hand–eye co-ordination.
- A practical activity may be classified either as an experiment or as an investigation.
- In an experiment the values of quantities are usually determined.
- In an investigation the effect of one variable upon another may be examined, or the effects of two similar physical conditions on a system may be examined or compared.
- Whether a practical activity is an experiment or an investigation, it must be planned and designed before being carried out. Once it is carried out, it should be reported on.
- The plan is a brief account of what is to be done.
- The design is a detailed account of how the plan will be carried out.
- The report is an accurate account of all that was done in carrying out the activity. It should include a list of the equipment used, how it was used, what readings were taken, how they were used, and how the information was used in arriving at a conclusion. The conclusion must always be based on the aim.
- All measurements are subject to errors. Some of these are due to limitations on the part of the experimenter – these are called random errors. Others are due to defects in the instruments used – these are called systematic errors.

- Random errors are due mainly to faulty reading of the scales of analogue instruments by the experimenter, and the faulty use of equipment.
- Systematic errors are usually zero errors in analogue instruments.
- To reduce the effect of random errors, we take an average of a large number of readings or, if a graph is drawn, we draw 'the best line' among the plotted points.
- To remove the effect of systematic errors in a reading, we either correct the defect or apply a zero correction.

Answers to ITQs

ITQ1

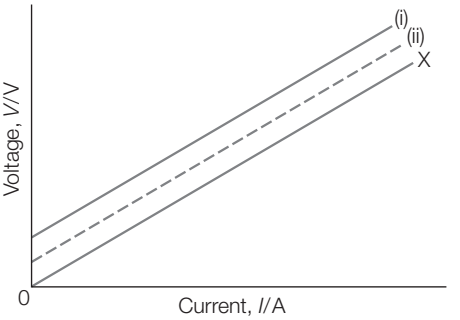


meter A (reads 0.5A) meter B (reads −0.3A)

ITQ2 Immerse the bulb of the faulty thermometer in melting ice until the reading is steady. Note this reading. The difference between this reading and 0.0°C is the zero error at 0°C.

ITQ3

Note: all three graphs must have the same slope. This slope represents the resistance of the resistor.



Examination-style questions

1 Which skills are required in doing each of the following? The first one is done for you.

(i) Using no-parallax to locate an image formed by a converging lens.

[Answer: Experimental skills: handling, manipulating, observing]

(ii) Correcting the reading of a micrometer screw gauge for zero error.

(iii) Measuring the length of a glass rod with a ruler.

(iv) Deciding whether to use a long pendulum or a short one to find a value for the acceleration due to gravity.

(v) Plotting the magnetic field of a bar magnet.

(vi) Calculating the slope of a graph.

(vii) Timing the oscillations of a simple pendulum.

(viii) Comparing the strengths of two bar magnets.

2 Your friend claims that a tennis ball will bounce higher when it is warm than it will when it is cold. Prepare a plan and a design for an investigation to decide whether your friend is right. This is a YES/NO investigation, so your design must include a hypothesis.

3 I have prepared a plan for an experiment to find the value for the acceleration of free fall (g) by allowing a pebble to fall from a tall building to the ground below. The building is about 30 m high. I have a tape 100 m long (× 0.5 cm) and a stopwatch reading to the nearest 0.01 s. The formula I hope to use to calculate the value of g is $h = \frac{1}{2}gt^2$

(i) State what readings I should take and show how they should be presented.

(ii) Explain how I should use these readings to reach a conclusion.

(iii) State what are the likely limitations of such an experiment.

(iv) How many significant figures can I justifiably use in stating my result for g? Give a reason for your answer.

4 An experiment to determine the focal length of a converging lens was carried out, in which image distances (v) were found for six object distances (u). The values of v for u increasing and for u decreasing are shown in the table below.

(i) What do the headings 'u increasing' and 'u decreasing' convey to the reader?

(ii) Copy and complete the table, expressing each value to an appropriate number of significant figures.

u/cm	v/cm		mean v/cm
	u increasing	u decreasing	
12.0	61.0	61.4	61.2
20.0	20.5	20.3	20.4
30.0	15.0	14.6	14.8
40.0	13.3	13.0	
50.0	12.5	12.6	
60.0	12.0	11.9	

(iii) Given a sheet of graph paper 19 cm × 15 cm, suggest suitable scales for u (horizontally) and v (vertically). (Hint: Consider the range of u values to be from 10 cm to 60 cm, and of v values to be from 10 cm to 65 cm.)

(iv) Draw a graph of v against u.

(v) On the same axes draw the graph of the equation $v = u$. The values of u and v at the point where the two graphs intersect are each equal to $2f$, where f represents the focal length of the lens.

(vi) Deduce the focal length of the lens.

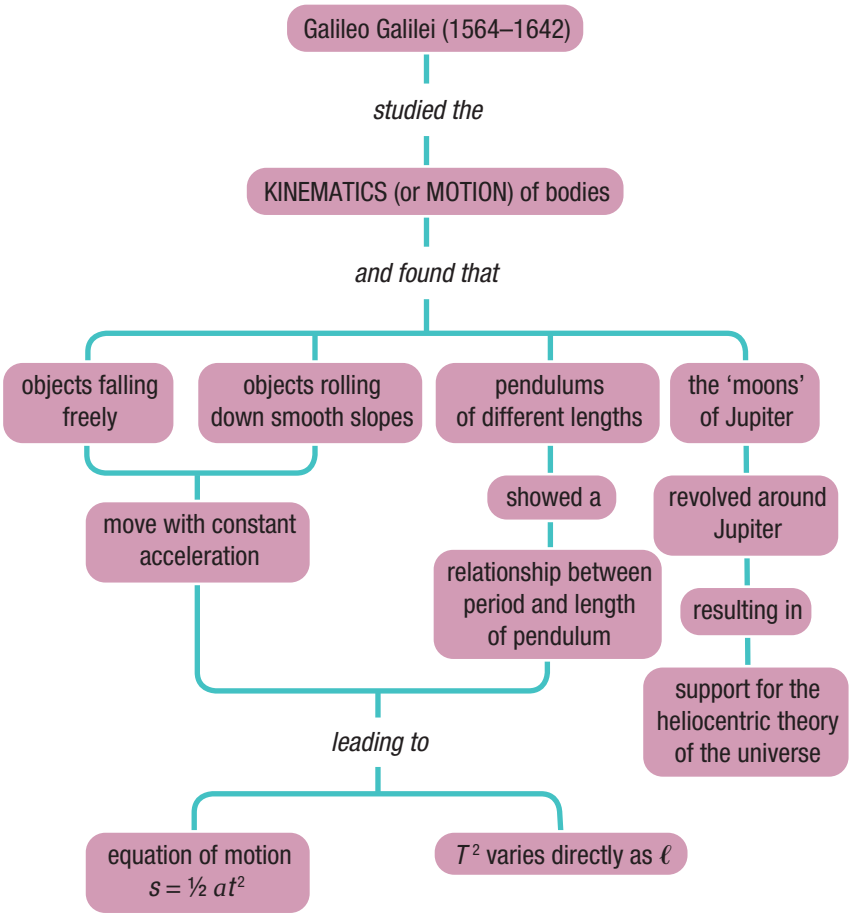
Section A:

Mechanics

4 Galileo Galilei and the Simple Pendulum

By the end of this chapter you should be able to:

- discuss Galileo Galilei's contribution to the methods of studying science
- understand why Galileo is regarded as the 'father of experimental science'
- explain why the simple pendulum is described as 'simple'
- understand the meanings of the terms associated with the simple pendulum
- appreciate the factors which determine the period of oscillation of a simple pendulum
- appreciate that straight line graphs can be used to determine how each of these factors affects the period of a simple pendulum
- understand why the pendulum was used in earlier times as the basis for constructing clocks



Galileo Galilei – what was his contribution to scientific methodology?



Figure 4.1 Galileo Galilei.

Galileo Galilei (1564–1642) was an Italian mathematician and astronomer who devoted most of his academic life to the study of motion. He was mostly concerned with what we now know as kinematics, which is the study of the features of motion, i.e. how acceleration, speed and distance are related to one another. Of particular interest to Galileo was finding out why some motions take place in a straight line, and why some, like the orbits of planets, do not. Explanation of this was only provided by Isaac Newton almost half a century after Galileo had died.

Galileo is known as 'the father of experimental science' because he was one of the first thinkers to test his hypotheses by carrying out practical experiments. Before he carried out his famous experiments at the leaning tower of Pisa and in the cathedral there, scientific ideas were formulated mostly on the basis of mere thought and neither proven nor discounted by physical proof. For example, centuries earlier, the Greek philosopher Aristotle propounded that it was force that determined the speed of a moving object, though he didn't carry out any experiments to prove it. So great was the respect for his views that almost everyone believed this theory without question. Aristotle's view therefore held sway for centuries until it was debunked by Newton in his famous work *The Principia* (1686).

It was Galileo who first formulated the equations of straight line motion by rolling spheres down smooth slopes and allowing spheres to fall vertically through large distances, taking careful measurements of distance and time to obtain the equations we now use. Galileo also discovered the relationship between the length of a pendulum and its period of oscillation. Using his pulse as a timer, he noticed that the time taken by a hanging lamp to complete one swing did not vary with the angle through which it swung. By measuring the time taken for pendulums of different lengths to complete their 'swing', Galileo was able to show that the square of the period was proportional to the length of the pendulum, or, stated in mathematical terms, that $T^2 \propto l$, where T is the period or the time for one swing to be completed and l is the length of the pendulum.

We shall examine whether these relationships are true or not by setting up our own investigations.

What is the relationship between the mass of the pendulum bob and the period of the pendulum?

Galileo, through his careful experiments, was able to test existing theories, and to cultivate and assess his own ideas. He even developed a telescope (known as 'the Galilean telescope') which he used to observe the movement of the Moon and Jupiter, allowing him to obtain evidence supporting his radical view that the Earth moved around the Sun (**'heliocentric' theory**). This new idea discounted the prevailing opposing 'geocentric theory' which advocated that planets moved around the Earth, a view that was advocated by the Catholic Church at that time. By holding and publishing the heliocentric view Galileo offended the authorities of the Catholic Church, and was found guilty of heresy by the Spanish Inquisition, a crime so severe that he could have been sentenced to death. However, in consideration of his old age, his punishment was commuted to house arrest.

'heliocentric' theory ›

Galileo died in 1642, the year Isaac Newton, the English mathematician and philosopher, was born.

Up to the early part of the last century many of the clocks in existence were 'grandfather' clocks where pendulums formed an important part of the structure. Although the pendulums used in these clocks were certainly not 'simple' in the sense explained on the next page, they were an integral part of the mechanism since their time of swing was found to be largely constant, meaning they kept relatively accurate time.

Grandfather clocks have since been superseded by clocks using flat coiled springs or still more recently by clocks using a quartz crystal (or by digital clocks), which are much more reliable and accurate.

What is a 'simple' pendulum?

We have all seen objects swinging to and fro, though not all of them could have been described as 'simple'. So what is a simple pendulum?

In our investigations that relate to the pendulum we will be using the following terms:

Suspension: This is the length of thin, light string or thread from which the heavy object will hang.

Bob: The object (commonly spherical) which hangs at the lower end of the suspension.

An oscillation: The movement between successive transits of the pendulum (bob and suspension) past a reference mark in the same direction.

Amplitude: The maximum displacement of the pendulum bob from its rest position.

Period: The time taken for the pendulum to carry out one complete oscillation.

Frequency: The number of oscillations the pendulum executes in one second.

Angular displacement: The angle which the suspension makes with its rest position at any moment.

A simple pendulum is an arrangement consisting of a heavy, but relatively small, mass hanging from one end of a length of light string or thread, with the other end of the thread or string attached to a fixed (i.e. immovable) support. If the string or thread used is not light, but thick and heavy and the mass not small compared with the thickness of the string, the pendulum is not considered 'simple'.

The rest of this chapter will be devoted to studying the physics of the simple pendulum – not the forces which cause it to swing from side to side, but the factors which affect the time that its swing takes. Most of us would know from experience that one factor affecting the period of a pendulum's swing is the length of the suspension. You might have noticed this in children's swings as well. There are other factors which also have an influence, such as:

(i) the mass of the body at the end of the suspension;

(ii) the size of the body;

(iii) the angle through which the pendulum swings.

Although these factors are not as effective as the length, they are, nevertheless, of sufficient importance to be worth investigating later in this chapter. We begin, however, with the effect of the length of the pendulum on the time of swing.

Practical activity 4.1

This activity is best done as a group effort (a group of perhaps five or six students) in which each member will have a turn at performing each of the many skills involved. These skills are:

(i) measuring the length of the pendulum;

(ii) determining the start and the end of an oscillation;

(iii) counting oscillations; and

(iv) timing oscillations.

The effect of the length of the pendulum on the period

How is the period of pendulum of a simple pendulum related to the length? This activity will test the relationship between the period and the length of the pendulum.

The aim

To investigate the relationship between the period, T, of the pendulum and its length, l, by plotting the period against the length and, if the relationship is not revealed by the graph of T against l (because it is not a linear relationship), to plot a graph of T^2 against l.

You will need:

- a length of thread or fine string about 160 cm long
- a small metal marble or, alternatively, a small heavy sphere (e.g. one made from a lump of plasticine), about 2 cm in diameter
- a fixed point of support (e.g. a split cork in the clamp of a tall retort stand with a weighted base) or, better, a metal hook screwed into the laboratory ceiling
- about 200 cm of strong twine to help measure the length of the pendulum if it is much longer than a metre
- stopwatch (digital or analogue)
- metre rule.

What are the variables? Which ones will be controlled?

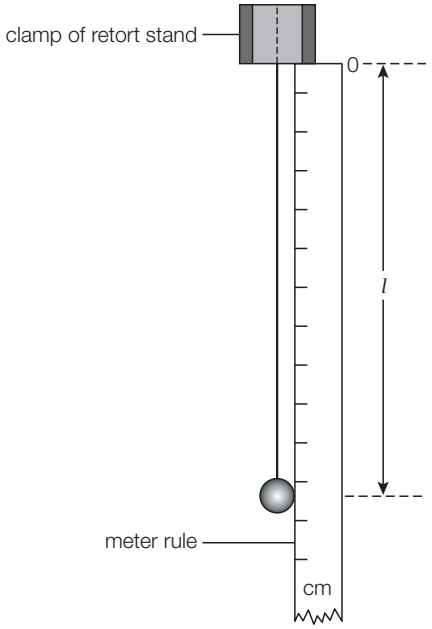
The factors which affect the period are the initial angular amplitude, the size of the bob, the mass of the bob and the length of the suspension. Since we are examining the effect of the length of the suspension, we will vary only this quantity and hold the others constant. This means that we will use the same initial amplitude throughout and also the same bob, and vary only the length of the suspension.

The variables to be controlled are (i) the mass of the bob; (ii) the radius of the bob; (iii) the initial angle of swing.

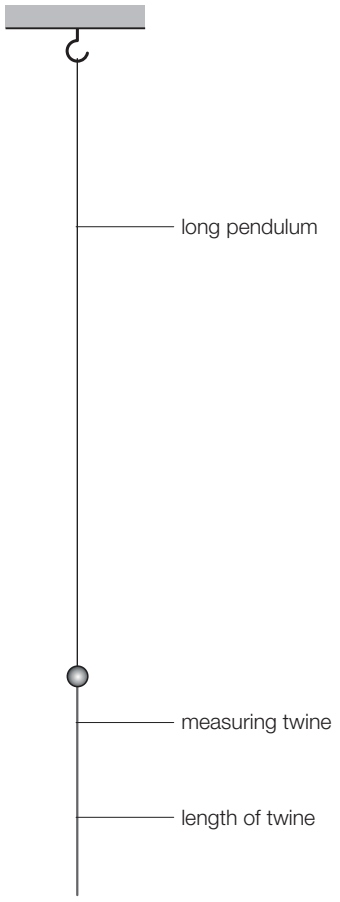
Method

1 Firmly attach the bob to one end of the length of string or twine.

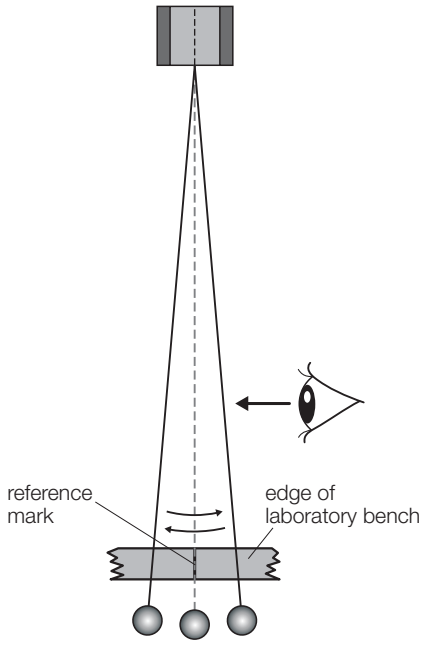
2 Using the metre rule, mark off (with ink) a length of 150.0 cm on the length of strong twine. Place the suspension string or thread in the slit of the cork and pull it through the slit to measure off a length of 150.0 cm between the split cork and the centre of the bob, using the length of twine already measured (see figure 4.2 (a) if a split cork is used or figure 4.2 (b) for a fixed



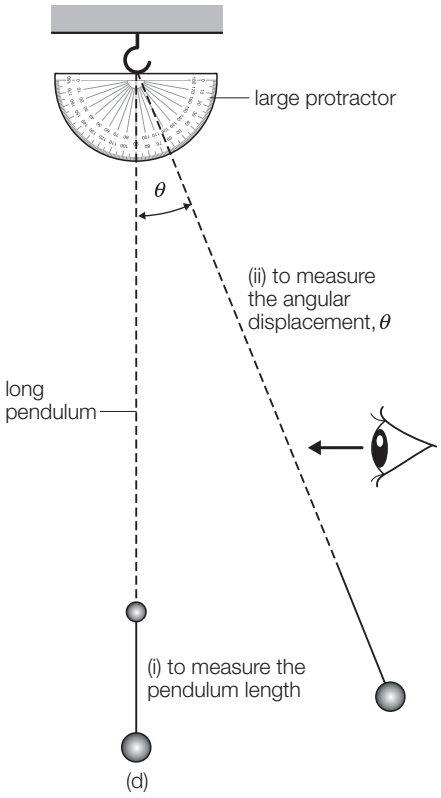
(a) short pendulum



(b) long pendulum



(c)



(d)

Figure 4.2 How to measure a pendulum. (a) To measure the length of the pendulum the graduated edge of the rule must be parallel to the suspension. The length is taken from 0 to a point on the rule where the bob touches the rule. (b) To measure the length of a long pendulum, a twine is stretched from the point of attachment to touch the bob tangentially. Marks are made at these two points and the distance between them measured against a metre rule. (c) For a short pendulum a reference mark is placed on the edge of the bench. (d) For a long pendulum the reference mark could be a pencil placed directly behind the rest position of the pendulum. The angular displacement is measured with a large protractor.

support). In doing this the measuring twine is held parallel to the stretched suspension. When the required length is obtained, tighten the clamp holding the cork to prevent slipping of the suspension as the pendulum swings or make fast the twine if a hook is used

3 On a table like that shown in table 4.1 record the length (l) of the pendulum in metres in the appropriate column. Allow the pendulum to come to rest so that the suspension hangs past (and close to) the edge of the laboratory bench.

Note that in counting oscillations you start counting from 0 and not from 1.

4 Make a chalk mark on the edge of the bench immediately behind the stationary suspension (figure 4.2 (c). This mark will be called a **reference point**. An alternative to the reference mark is the reference point, which could be a pencil held in a horizontal position in a retort stand, immediately behind the stationary suspension. Whether you use a mark or a point, it will be used to judge the beginning and the end of oscillations.

reference mark (or reference point) ›

5 With the suspension pulled taut, displace the bob through a small distance (about 2 cm) to one side, release it and closely examine the movement of the bob from the side to see whether it is moving along an arc and not along an ellipse (see figure 4.2 (c) or (d)). This is an important precaution, if the theory we are assuming is to apply.

Note that we are taking care to change only the length of the pendulum. All the other factors which might affect the period, e.g. size of the bob, mass of the bob and the initial displacement of the bob, will remain constant. We are 'controlling' those variables and changing only the one, length, whose effect we are examining.

6 With the pendulum swinging in the correct way and your thumb on the stopwatch start button, stand in front of the stationary suspension and decide whether you will regard one oscillation as the movement between successive transits of the bob to the left or to the right (whichever you prefer) past the reference mark on the bench or the reference point provided by the pencil.

7 With the pendulum swinging, begin counting from –5 up to zero and, just as the bob is moving to the left (or right) past the reference mark (or point), start the stopwatch and begin counting the oscillations from 0, going upwards from 0, 1, 2, 3 and so on up to 10.

8 Just as you say 'ten' and the bob is making its tenth (10th) transit to the left (or right) past the reference mark (or point), stop the stopwatch. Read and record on the table the time interval between the starting and stopping of the watch (see table 4.1). This is the time for 10 oscillations of the pendulum.

9 Repeat steps 1 to 7 above for the shortest proposed length, l, of the pendulum (= 0.1000 m) (why 4 significant figures? What does this tell the reader?)

10 Having decided that the range of l will be from 150.0 cm to 100.0 cm, enter the values for T and T^2 on the table for each of these lengths.

11 Set up axes for T/s (as ordinate, that is, along the y-axis) and l/m (as abscissa, along the x-axis) with suitable scales, each starting from 0, and plot the two points corresponding to the two pairs of values of T and l so far obtained.

12 Carry out steps 1 to 7 and plot the corresponding points for l = 110.0 cm, 120.0 cm, 130.0 cm, 140.0 cm, in turn.

Table 4.1 Table of readings.

Observation	Length of pendulum, l/m	Time for 10 oscillations, T_{10}/s	Time for one oscillation, T/s	T^2/s^2
1				
2				
3				
4				
5				
6				
7				

In entering your values on the table:

(i) Check that you have an appropriate degree of significance for all the values on each column. If your measurements were made to the nearest millimetre, what will be the entry for l when its value is 100 cm?

(ii) Check that the degree of significance of all the readings in each column is the same.

The degree of significance is not expected to vary from reading to reading in any one column. Can you say why?

The graph

Now use the values you obtained from your experiment to plot the graph of:

(a) mean T against l, and

(b) (mean T)^2 against l.

Now discuss the way the activity went. You should make reference to some of the points raised below.

If your graph (i) turns out to be a curve, it will be difficult to determine its equation. All that you will be able to say might be that as l increases, the period, T, increases, but we could not say exactly how. You would then plot the graph of T^2 against l and find that the graph was straight and passing through the origin. Its equation would then be

$T^2 = S \times l$, where S is the slope, a constant.

This is an example of 'straight line graph theory' being applied to the pendulum.

This would imply that

T^2 is proportional to l or $T^2 \propto l$

Calculate the slope of your graph. Note that the unit of this slope will be the units of T^2 divided by the units of l, that is, s^2/m or s^2 m^{-1}.

In this book both 'slope' and 'gradient' are used to mean 'change of y value (with units)' divided by 'change of x value (with units)'. This ratio generally will have units which must be stated with the value of the slope.

Now calculate the slope of graph (b) (do not forget the units!). Express this slope to an appropriate number of significant figures. How many?

The value of the slope of graph (b) should be 4.03 s^2 m^{-1}. What is the difference between this value (the 'expected' value) and the value you obtained from your graph (a)? Calculate the percentage of the 'expected' value which this difference represents.

Taking this percentage difference into consideration, would you say that your experiment was satisfactory?

Remember that the success or otherwise of your result will have depended on the precautions taken in the measurements of l and T.

Possible difficulties and limitations

Any pendulum activity carried out in an ordinary laboratory will be subject to difficulties of many different kinds, such as the following:

(i) The presence of a draught in the laboratory (in the Caribbean where windows are kept open as a rule). These will cause the swings to be irregular and somewhat unruly!

(ii) The possibility that the bob may not move along an arc, or the suspension, in one plane.

(iii) The possibility of the suspension slipping through the split in the cork or the point of support moving as the

oscillation proceeds – leading to l changing as swinging proceeds.

(iv) The difficulty of determining the start and the end of oscillations – uncertainty in the values of T.

(v) The point of support may not remain absolutely fixed during oscillations.

Steps that can be taken to reduce the effect of these difficulties are:

(i) conduct the activity in the far corner of the laboratory, far away from the windows;

(ii) exercise patience and wait for the pendulum to swing satisfactorily (as described above);

(iii) ensure that the clamp holding the slit cork is properly tightened, and, if tied instead to a hook, that the knot made is secure and so cannot become undone;

(iv) use a reference mark and to ensure that the pendulum is moving past this mark in the same direction at the start and at the end of an oscillation;

(v) use a heavy mass on the base of the retort stand if a split cork is used to hold the suspension.

It is best not to push the bob to start the oscillation, but just to release it when the desired initial angular displacement is obtained.

ITQ1

Table 4.2 shows some specimen values for T^2 and l for a pendulum experiment. Taking into consideration possible reaction error in timing the oscillations and the calibration of the metre rule, and referring to table 4.2, discuss the following:

(a) Was the use of 4 significant figures for l in table 4.2 justified?

(b) Was reaction error taken into account in fixing the significance of T^2?

ITQ2

Use the values shown in table 4.2 to plot the graphs of (i) T against $l^{½}$ and (ii) T^2 against l.

Calculate the slope of each graph. Is the ratio of the slopes of two graphs what it is expected to be from theory, which states that $T = 2\pi\sqrt{l/g}$?

MATHEMATICS: algebra (simultaneous equations)

Table 4.2 Specimen values for T^2 and l for a pendulum experiment.

Observation	Length of pendulum, l/m	Time for 10 oscillations, T_{10}/s	Time for 1 oscillation, period, T/s	(Period)^2, T^2/s^2
1	1.500	24.5	2.45	6.00
2	1.400	23.8	2.38	5.66
3	1.300	22.9	2.29	5.24
4	1.200	22.0	2.20	4.84
5	1.100	21.1	2.11	4.45
6	1.000	20.0	2.00	4.00

Worked example 4.1

A pendulum of length l_0 metres with a small angle of swing has a period of 2.00 seconds. What would be the period of the pendulum if its length were (i) halved, and (ii) doubled?

Solution – first method

Use the relation $T^2 = S \times l$, where S is a constant. Assume the first time is T_1 and the new time is T_2. Also assume that the first length is l_1 and the new length is $l_1/2$. We have, on substituting for the two cases,

$T_1^2 = S \times l_1$ (equation 1) and $T_2^2 = S \times \frac{l_1}{2}$ (equation 2)

Divide equation 1 by equation 2; we have $\frac{T_1^2}{T_2^2} = \frac{l_1}{(l_1/2)} = 2$.

So $T_2^2 = \frac{1}{2}T_1^2$, giving T_2, the new time of swing = $\sqrt{½}$ of the old time of swing.

So the new time of swing will be $0.71 \times 2.00\,\text{s} = 1.41\,\text{s}$, since $\sqrt{½} = 0.707$.

MATHEMATICS: algebra (direct variation)

ITQ3

Taking constant $S = 4.03\,s^2\,m^{-1}$ in the equation $T^2 = S \times l$ calculate the length of a simple pendulum whose half-period ('tick-tock') is 1 second.

Now attempt a solution using the method of simultaneous equations.

Solution – alternative method

By the variation, if T^2 varies as l. Since l is *reduced* by a factor of 2, then T^2 will also be reduced by a factor of 2, and T must be reduced by a factor of $\sqrt{2}$, i.e. by a factor of 1.41, the same result as before. The new time will therefore be 1/1.41 of the old time, or $0.707 \times 2 = 1.41\,s$.

Solution to (ii) (using the variation method)

(ii) If the pendulum length were doubled, then T^2 would be doubled, and thus T would increase by $\sqrt{2}$ (or by 1.41). If the original period was 2.00 s, the new period would now be $\sqrt{2} \times 2.00\,s$ or 2.82 s.

Practical activity 4.2

Does the initial displacement of the pendulum bob (or the initial amplitude) affect the period of the pendulum?

We must treat this activity as an investigation in which we are attempting to find the answer to the question posed above. What is the aim of the investigation?

Aim

To find out whether the size of the initial displacement of the pendulum has any effect on the period of the pendulum and, if it does, what is that effect?

Since this is an investigation and there is only one answer to the question posed, the answer to the question is either YES or NO, and so we can express an opinion. We will say 'YES' to the question and state our hypothesis as follows:

Hypothesis (part 1)

We think that the initial displacement of the pendulum does have an effect on the period of the pendulum.

Hypotheses (part 2)

We think that the period of the pendulum will increase as the initial displacement increases. This hypothesis is based on the experience gained from children's swings. Swings seem to take longer to do one oscillation when the angle of swing is large.

What is the plan?

We will vary the initial angle made by the suspension of a long pendulum with the vertical over a wide range of values and note the average period of the pendulum for each angle used. A graph of period against the initial angular displacement will make clear how the two variables are related.

You will need:

- about 2.3 m of thread
- length of twine about 3 m
- a plumb-line (on a retort stand) to determine verticality and to serve also as a reference mark for the oscillations
- large protractor
- stopwatch
- retort stand with a split cork held in the clamp or (better) a fixed support (like a hook in the laboratory ceiling) to attach the pendulum
- metal or plasticine bob about 2 cm in diameter.

What are the variables? Which ones will be controlled?

The variables to be controlled are (i) the mass of the bob; (ii) radius of the bob; (iii) the length of the suspension. We will vary the initial angle of swing.

The two variables will be the initial angular amplitude, θ, (the independent variable) and the corresponding period of swing, T, (the dependent variable). We will use values of θ from 10° to 70° at intervals of 10°.

Method

1 Attach the bob securely to the length of thread.

2 Place the length of thread in the split of the cork and pull it through the cork

until the length between the point where it leaves the cork and the centre of the bob is about 2.00 m as measured with the metre rule. Tighten the clamp to ensure that the thread does not slip through the cork as swinging proceeds. Alternatively, if a hook is used as fixed support, pass the free end of the suspension over the hook in the ceiling and adjust the length of thread from hook to bob (by pulling the thread) to about 2.00 m. Now tie the suspension securely to the hook at this length.

3 Set up the plumb-line directly behind the stationary pendulum.

4 With the suspension pulled taut, incline it to an angle of 10° with the plumb-line, using the large protractor to measure the angle (see figure 4.2 (d)). Release the bob.

5 From the side (figure 4.2 (d)) check to make sure that the bob is swinging along an arc and not in an ellipse.

6 As explained in Practical activity 4.1 on page 54, measure the time for 10 oscillations and record this time, T_{10}, as well as the corresponding angle, θ, on a table like that shown in table 4.3. Calculate and record the time, T, for one oscillation.

7 Repeat steps 4 to 6 for the initial angular displacement equal to 70°.

8 Now that the values of the variables for $\theta = 10°$ and $\theta = 70°$ have been determined, set up axes using suitable scales for θ and T in order to ensure a large graph.

9 Repeat steps 4 to 6 for the remaining angles of 20°, 30°, 40°, 50°, 60°.

10 Plot a graph of the mean of T against θ.

Table 4.3 Table of results.

Initial angle of swing, θ/°	Time for 10 swings, T_{10}/s				Time for one swing
	(i)	(ii)	(iii)	mean	T/s
10					
20					
30					
40					
50					
60					
70					

Use of the graph

The shape of the graph will determine the decision to be taken and a conclusion drawn as to whether the initial amplitude of the swing has an effect on the period of swing. If the graph is parallel to the θ-axis when the best line is drawn among the points, it will imply that regardless of the value of the initial angle of swing, the time of swing remains constant. Any other shape would suggest that the period of the pendulum changes with the initial angular displacement. If the graph curves upwards as θ increases, the decision will be that the period increases with the initial angle of displacement. From our hypothesis we expect this to be the case. If the curve slopes downwards then the opposite will be the case. The expectation is that as θ tends to 0°, the graph is expected to become more and more nearly parallel to the θ-axis.

Possible difficulties and limitations

As well as the difficulties and limitations mentioned before in Practical activity 4.1, there is a further difficulty in this activity. As the oscillations proceed the amplitude slowly reduces because friction from the atmosphere will retard the motion of the pendulum. To say, therefore, that the oscillation has taken place at a constant initial amplitude would be false. This is a true limitation. We have no control over the cause.

Now write a discussion of how your activity went, using the 'pointers' given before and, after this, draw a conclusion, remembering that the conclusion must be related to the stated aim.

ITQ4

What steps can we take to reduce the effect of the limitation mentioned at the end of Practical activity 4.2?

Questions for class discussion

1 Why is the chosen length of the pendulum so large?

2 What are the forces acting on the bob as it moves?

3 Is the tension in the suspension one of the forces affecting the acceleration of the bob? If not, why not?

Practical activity 4.3

What are the advantages of using a long pendulum like that used in Practical activity 4.2?

In all the pendulum experiments we have discussed we have had to measure the time interval over which 10 oscillations of the pendulum occurred. As discussed in chapter 2, we must acknowledge the constant presence of a reaction error in all of these measurements and since this error could be an appreciable fraction of the period of a short pendulum, we have deliberately decided to use long pendulums which will have periods that are much larger than possible reaction times. Moreover, by measuring the times for 10 swings, we will spread the reaction error over a number of swings thereby ensuring that the percentage error in a swing is further reduced.

Does the mass of the bob affect the period of the pendulum?

This activity is very similar to Practical activity 4.2, except that the mass of the bob will be altered; the length of the suspension, the dimensions of the bob and the initial angle of swing will remain constant. While in the last activity we could have used any angle we wished, we cannot use any mass we wish in this one, unless we are prepared to create these masses ourselves. With a bit of serious thought we can work out how this can be done, remembering that the dimensions of the bob must be held constant and be very small compared with the length of the pendulum.

We will use a small plastic cup with a tiny hole at the centre of the base of the cup through which we will pass the thread of the suspension. The end of the thread that is passed through the cup is tied on to a matchstick below the cup.

Again, as before, there is a 'YES or NO' answer to the question as to whether the mass affects the period and so we can state an aim and formulate a hypothesis.

The aim

To find out whether the mass of the bob of a simple pendulum affects the period of oscillation of the pendulum.

The hypothesis

This may be stated as follows: We feel that the mass of the bob of a pendulum does not affect the period of the pendulum, provided that the bob is of small size and is not seriously affected by frictional forces due to the atmosphere.

You will need:

- 3.20 m length of strong thread for the suspension
- 250 ml plastic cup with a pinhole at the centre of the bottom
- a matchstick
- a stopwatch
- pencil in a retort stand to serve as a reference point for the oscillations
- 0.5 kg of sand in a beaker
- hook in a beam in the laboratory ceiling to serve as the point of attachment of the pendulum.

Method

1. Pass one end of the thread through the pinhole in the bottom of the cup. Tie this end of the thread on to the matchstick.
2. Attach the other end of the thread to the hook in the laboratory ceiling.
3. Pour 50 g of sand into the plastic cup. Level the sand in the cup.
4. Position the reference pencil immediately behind the stationary suspension.
5. With the suspension taut, displace the plastic cup containing the sand about 5 cm from its mean position, release it and observe the oscillation from the side (see figure 4.2 (d)).
6. When the movement of the cup is along an arc and the movement of the suspension is in a plane, begin to observe the oscillations from a position directly in front of the reference mark.
7. Measure the time for 10 oscillations of the pendulum and record it on a table similar to table 4.3, but with the first column containing the masses of the bob that are used. Disregard the mass of the cup.
8. Repeat this measurement twice more and record the results on the table.
9. Repeat steps 5 to 8 with the mass of sand in the cup equal to 400 g.
10. Having calculated the time for one swing for each of the masses of 50 g and 400 g, set up axes, using scales for the time for one oscillation and the mass of the bob, which will result in a large graph.
11. Plot the points corresponding to these two pairs of co-ordinates.
12. Now repeat steps 5 to 8 for the mass of sand in the cup equal to 350 g, 300 g, 250 g, 200 g, 150 g and 100 g in turn, enter the relevant points on the table and draw the line of best fit among the points.

ITQ5
Calculate the length of a simple pendulum which has a period equal to a reaction time of 0.2 seconds.

ITQ6
Why is it a good idea to time a large number of oscillations if the pendulum length is small? Why should the number of oscillations timed be increased as the length of the pendulum decreases?

Use of the graph

You will now be able to determine the outcome of the investigation from the shape of the graph you obtain. From the hypothesis formulated above we expect the graph to be parallel to the mass axis. So if the best line is parallel to the mass axis, then the hypothesis will have been upheld. If not, the implication will be that the mass of the bob does have an effect on the period of the pendulum. What this effect is will, of course, depend on the shape of the graph obtained.

Limitations and difficulties

These will be the same as those mentioned above, except that the effect of any draughts there might be will be less severe than they were before because of the larger masses used in this investigation. At the same time, however, since the surface area of the bob in this investigation would be larger than before, there will be a greater 'resistance' caused by the air (called 'drag') and this would have a 'slowing down' effect on the oscillation.

Now write a discussion along the lines of previous discussions and end with a conclusion.

Chapter summary

- Galileo was interested more in 'how' bodies moved (called kinematics) than in 'why' they moved as they did. He showed that for a pendulum of length l oscillating with period T, T^2 varies directly as l.
- Galileo developed his own version of a telescope, called the Galilean telescope, by means of which he was able to view the moon's surface and the planets of Jupiter.
- Galileo became an adherent of the 'heliocentric' theory of the universe which advocated that planets revolved around the Sun and not the other way round (called the 'geocentric' theory).
- Galileo was placed under house arrest by the Catholic Inquisition because his view was considered heretical.
- The factors that affect the period of a simple pendulum are:
  (i) the length of the pendulum;
  (ii) the size of the bob of the pendulum;
  (iii) the initial angular displacement of the pendulum.

Answers to ITQs

ITQ1 **(i)** Yes, since precision of measurement of l = 1mm and distances are given to the nearest 0.001m or the nearest mm.
(ii) Total reaction error over 10 oscillations is about ± 0.4s. For one oscillation, reaction error is 0.04s. Values of timing show an error per oscillation as 0.01s. Since this is much smaller than 0.04s, we must presume that reaction error was not taken into account. Since this would be the precision if reaction error was considered and it is much larger than the 0.01s, shown by the tabulated values, we must presume that reaction error was not considered.

ITQ2 Slope 1 = $2.02\,s\,m^{-0.5}$; slope 2 = $4.05\,s^2\,m^{-1}$
So $\frac{(\text{slope 1})}{(\text{slope 2})} = 0.499\,s^{-1}\,m^{0.5}$
The theoretical value of this ratio = $0.502\,s^{-1}\,m^{0.5}$

ITQ3 Length of pendulum = 99.2cm

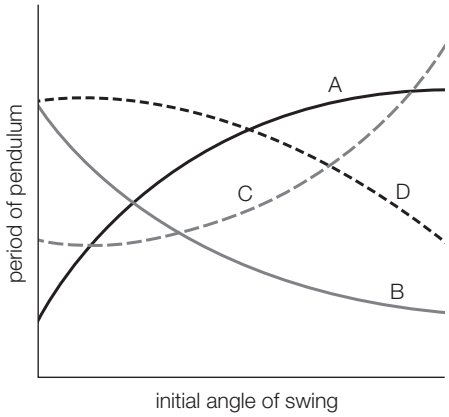
ITQ4 Use a long pendulum (large period) with a small, but heavy, bob (smaller drag effect). Large period will produce a small fractional change of period; smaller drag effect will result in smaller changes in amplitude.

ITQ5 9.92 mm

ITQ6 Since $T^2 \propto l$, as l decreases T decreases and the reaction error becomes increasingly important to the period. To counter this, we should 'spread the error' over an increasing number of oscillations so as to make the reaction error per oscillation reduce as the period reduces.

Examination-style questions

1 A class of students is told that the period of oscillation of a pendulum increases as the initial angle of swing increases, but for small angles of swing the period may be taken to be constant. Which of the graphs A–D best represents this statement?



2 There are three swings in a children's playground. The effective lengths of the swings are 2 m, 3 m and 4 m respectively. Assuming that the square of the period of the oscillation of each swing is proportional to its length and that the period of the 2 m length swing is 2.8 s, calculate the period of each of the other two swings.

3 A student uses a long pendulum obtain a value for g, the acceleration due to gravity. He is well aware of the fact that, in order to obtain a satisfactory result, he must use as small an angle of swing as possible. His first determination is made by displacing the pendulum bob a horizontal distance of 5 cm from its equilibrium position and, by using a sine calculation, obtains an angle of swing of 2°.

(i) Draw a diagram showing:

- (by a full line) the length of the pendulum;
- (by a dashed line) the pendulum with the bob displaced 15 cm from its rest position;
- (by a dotted line) the displacement of 15 cm.

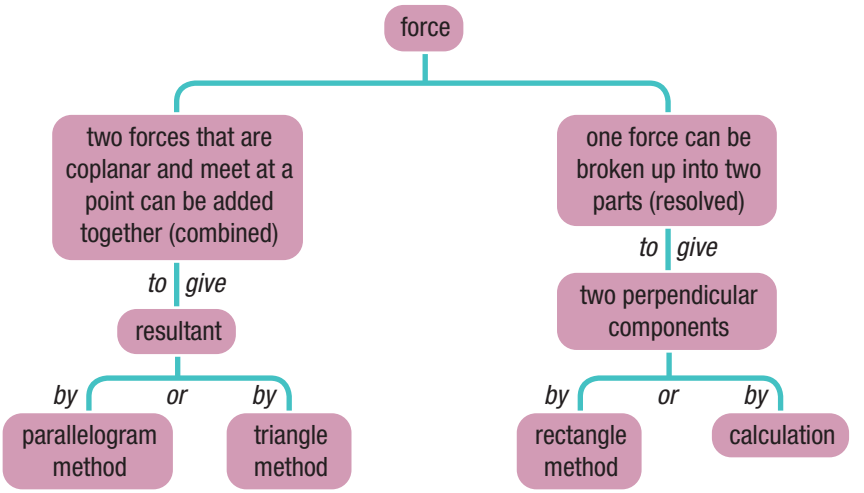
(ii) Calculate the length of the pendulum.

(iii) Calculate the value of g the student would have obtained from the experiment, taking the value of the constant S in the equation $T^2 = S \times l$ to be 4.03 $s^2 m^{-1}$.

5 Combining and Resolving Vectors

By the end of this chapter you should be able to:

- distinguish between a scalar quantity (or 'a scalar') and a vector quantity (or 'a vector')
- understand what is meant by the 'resultant' of two co-planar vectors
- add together (or 'combine') two vectors to find their resultant in magnitude and direction by using either a construction method, or a calculation
- resolve a vector into two parts (or 'components'), each acting perpendicular to the other



Scalars and vectors

scalar

vector

In your physics course you will be dealing with two types of quantities, namely scalars (or scalar quantities) and vectors (or vector quantities). Scalar quantities are those which have only magnitude (or size). Examples of **scalars** are mass, age and time. These clearly have no direction.

Vector quantities are those which have both magnitude (or size) and direction. Their effect depends on the direction in which they act. Examples of **vectors** are velocity, acceleration and force.

Thus the mass of a body may be 2 kg. There is no question of the direction in which this mass acts. If the age of someone is 14 years, there is no question of the direction of this age. So this is the nature of scalars.

Consider for a moment the high velocity winds of a hurricane. At any given moment the velocity of the wind will have a definite magnitude and be in a definite direction at a given place. The velocity at a different place may have the same value, but it may have a different direction. If it does, the wind there will have a different velocity. So velocity is a vector. My weight and yours will always act vertically downwards and each will have a definite value. My weight and yours are therefore vectors.

So, in physics, mass and weight are not the same type of quantity. One, mass, is a scalar, whereas the other, weight, is a vector.

In this chapter, we will be concerned with the difference between scalar quantities and vector quantities.

There are times when, in dealing with vectors (like forces and velocities) and their effects, we need to 'manipulate' them to suit the case we are dealing with. It may be good to bring the forces or the velocities together ('combine' them) in a certain case, and to break them up into smaller parts ('resolve' them) in another.

The quantities we deal with in physics, as we have seen, are of two quite distinct types and the way in which they are treated in calculations are quite different. One set of these quantities obey the simple laws of arithmetic and can be added and subtracted as in ordinary arithmetic; the other cannot.

Scalars: how they are treated

Quantities which conform to the ordinary laws of arithmetic are called scalar quantities or scalars. They are quantified only by their size. For instance, all I need to say about a loaf of bread to describe its physical quantity is to mention its mass (what we commonly, but wrongly, refer to as its 'weight').

Many of the quantities we meet and talk about in our ordinary everyday experience are scalars. Examples of these are volume, speed, mass, areas, age, sums of money, etc. As we know, such quantities can be added or subtracted in the usual manner.

If you give me 20 dollars now and 15 dollars 20 minutes later, I will always have 35 dollars altogether. There can only be one answer to the question of how much money I now have. My resultant sum is 20 dollars + 15 dollars, or 35 dollars. This is so only because a sum of money is a scalar quantity.

There is no question of money having any other attribute apart from its monetary value. It clearly has no direction associated with it. My final position, which I will call my resultant position, is that I now have 35 dollars. Let us now take a very different case, the case of a vector quantity.

Vectors: how they are treated

Imagine I am at a position O in a field. I walk 20 metres due east to a position X. My new position is different from my old position. My change of position is obviously 20 m but, in order to be specific, I must say 20 m due east of my former position. This change of position is called my **displacement** and it is represented by a line bearing an arrow and pointing to the east. This quantity, the displacement, has both magnitude (20 m), and direction (eastward). This is the nature of a vector quantity or a vector (see figure 5.1 (a)). It has both size ('magnitude') and direction. It is only completely defined if both its magnitude and its direction are stated. So displacement is a vector quantity.

displacement ›

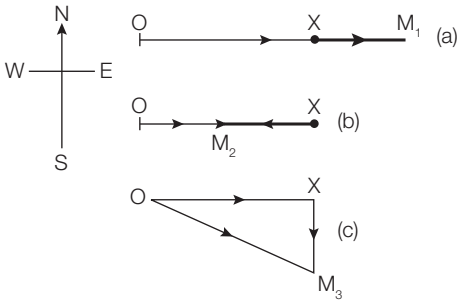


Figure 5.1

Let's now examine three different cases:

(i) I first walk 20 m due east and then I walk 15 m from X in a direction due east (figure 5.1 (a)).

(ii) I first walk 20 m due east and then I walk 15 m from X in a direction due west (figure 5.1 (b)).

(iii) I first walk 20 m due east and then I walk 15 m due south from the point X (figure 5.1 (c)).

The three additional displacements are all different displacements since, although they all have the same magnitude, 15 m, they each have a different direction. When I add these different displacements to the original displacement, I get three different **resultant** displacements.

resultant ›

In case (i) (figure 5.1(a)), where I first walk 20 m due east to X and then 15 m due east from X to a point M_1, I end up at M_1, 35 m from O. My new displacement or my resultant displacement is now how far I am from O and in

what direction. Clearly my resultant displacement is now 35 m from O in an easterly direction. It is given by the vector $\overrightarrow{OM_1}$.

In case (ii) (figure 5.1 (b)), I first walk 20 m east and then 15 m due west from X. I end up at a point M_2, 5 m to the east of O, and so my new displacement is 5 m due east of O and is given by the vector $\overrightarrow{OM_2}$.

In case (iii) (figure 5.1 (c)), where I first walk 20 m from O to the point X and then 15 m from X due south, I will end up at the point M_3, and my overall displacement will be $\overrightarrow{OM_3}$, whose distance from O is, by the geometry of the figure, 25 m, since

$$OM_3^2 = OX^2 + XM_3^2$$

giving
$$OM = \sqrt{(OX^2 + M_3X^2)}$$
$$= \sqrt{(20^2 + 15^2)}$$
$$= \sqrt{625}$$
$$= 25$$

So the resultant obtained by adding vector $\overrightarrow{XM}$ to vector $\overrightarrow{OX}$ (or combining them) depends completely on the direction of vector $\overrightarrow{XM}$ and since this vector $\overrightarrow{XM}$ can take an innumerable number of directions between due east and due west through south and, of course, through north as well, there will be an equally innumerable number of resultants that are all different in magnitude or direction or both magnitude and direction. A careful look at the situation will reveal that the maximum value of the resultant is obtained by adding the first easterly displacement of 15 m, which gives a resultant of 35 m due east and the minimum a resultant of 5 m by subtracting the second westerly displacement of 15 m or, alternatively, by adding –15 m to the original 20 m east. We could summarise all of this by saying, in mathematical language, that the resultant displacement, d, is given by $5\text{ m} \leq d \leq 35\text{ m}$.

Notation for vectors

When using letters (for example, $\overrightarrow{OX}$) to represent a vector, sometimes we draw an arrow above the letters to indicate the direction of the vector. In this case we could write $\overrightarrow{OX}$. The direction is in the order of the letters.

MATHEMATICS: algebra: the range of a variable, d

You can see why I can get a range of very different answers for my distance from my starting point. In figure 5.1 (a) d is clearly equal to 35 m and in (b) it is clearly 5 m. If I walk due south, **Pythagoras' theorem** gives d as 25 m.

Pythagoras' theorem

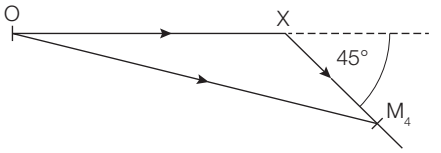


Figure 5.2 Using the triangle method to find the resultant of two vectors.

For any other direction I can use a scale diagram to obtain the answer. Thus, if instead of walking south as in case (c) above, I walk south east from X, I can draw a scale diagram to find my resultant displacement (see figure 5.2). I will use a scale of 2 cm to represent 5 m. I first draw a line OX 8 cm long to represent 20 m of actual distance, then at X construct an angle of 45° with the easterly direction in a clockwise direction and from it cut off a length XM_4 of 6 cm to represent 15 m. If we now measure the length OM_4 it will represent my displacement from O.

To get the result we have used what is known as the **triangle method** to find the resultant of two vectors, where the vectors added were $\overrightarrow{OX}$ and $\overrightarrow{XM_4}$. This method is based on the triangle rule for finding the resultant of any two vectors which are in the same plane (we say that they are '**co-planar**'). This rule states that:

the triangle method

co-planar

> **The resultant of two co-planar vectors may be represented by the third side of a triangle, two of whose sides are parallel to and proportional to the vectors and are constructed 'head to tail'.**

In using the rule, therefore, to obtain the answer to the case (i) above, you would:

1 Draw one side of the triangle, OX, to a scale of, say, 2 cm to 5 N.

2 From the 'tail' of OX (which is the point X) construct the second vector to scale; and so you would draw a line 6 cm from X in the same direction as OX. Denote the end of this vector by M.

3 Draw the third 'side' of the triangle by joining the origin, O, to the 'tail' of the last vector drawn. This vector represents the resultant of the first two vectors.

4 Measure the length of this third 'side' OM of the triangle and use the scale to determine the value of the resultant (see figure 5.2).

This is the method by which vectors are added using the triangle of vectors. In this case the vectors we added were displacement vectors, the term displacement being used because it represented the distance we moved from the starting point. So my first displacement from O was $\overrightarrow{OX}$, then my next displacement from X was $\overrightarrow{XM}$. My overall displacement, called the resultant displacement, is measured from the original starting point, O, to the point where I ended my journey.

'Collinear' means along the same line.

vector equation

We could represent this movement from O to M using a general **vector equation**:

$$\overrightarrow{OX} + \overrightarrow{XM} = \overrightarrow{OM}$$



Figure 5.3 The resultant of a vector $\overrightarrow{OX}$ and collinear vector $\overrightarrow{XM}$ is obtained by joining the origin O to the tail, M, of the last vector drawn. The resultant is $\overrightarrow{OM}$.

In a vector equation we give a positive sign to vectors directed to the right in accordance with convention. In keeping with this convention, vectors directed to the left are given a negative sign. Using this convention, the equation from figure 5.1 (b) would be:

$$\overrightarrow{OX} + (-\overrightarrow{XM}_2) = \overrightarrow{OM}_2$$

or $$\overrightarrow{OX} - \overrightarrow{XM}_2 = \overrightarrow{OM}_2$$

The treatment applied to displacement, a vector, may be applied to any other vector. Other examples of vectors you will meet in the 'mechanics' part of your course are velocity, acceleration, force and momentum. As we would expect for any vector, the effect produced by any one of these depends very much on the direction in which it is acting. Note again that in the examples just discussed, although there can be any number of displacement resultants, there can be only one value of **distance covered**, and that is 35 m. This is the value of the scalar quantity – distance covered.

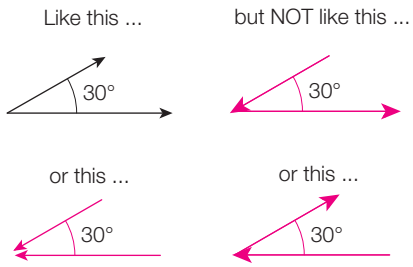


Figure 5.4

Angles between vectors – a word of caution

The angle between two vectors is the angle made when they are drawn with their direction arrows both pointing away from or both pointing towards the point where they meet. For example, two vectors with an angle between them of 30° should be drawn as in figure 5.4

distance covered

The parallelogram method for combining vectors

There is another method that can be used to add or combine two vectors. It is called the parallelogram method. This method is based on the parallelogram rule, which states that:

> **The resultant of two co-planar vectors is given by the diagonal, drawn from the origin, of a parallelogram, two adjacent sides of which are proportional to and parallel to the forces (or which have the same included angle as the vectors).**

Remember that the opposite sides of a parallelogram (figure 5.5) are equal in length.

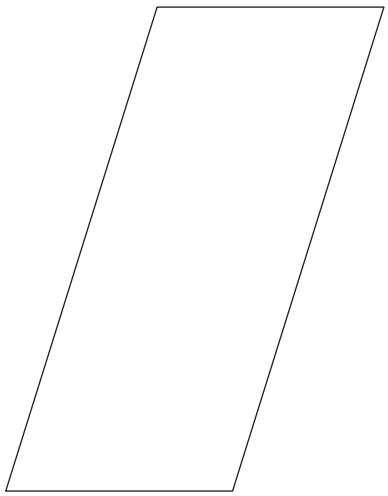


Figure 5.5 A parallelogram.

Whereas in using the triangle rule the vectors are added 'head to tail', where the head of the second vector is added to the tail of the first, in the parallelogram method either the heads or the tails of the two vectors are put together to start the addition. It is possible to use this method only if the two vectors are not acting along the same line.

For example, we could have added the displacement of 20 m directed due east in the example of case (iii) above to the 15 m directed due south by putting the head of the 20 m and the head of the 15 m together in doing the construction and then drawing the vectors in their respective directions. The diagram of figure 5.6 shows the procedure. We use the same scale of 2 cm to 5 N as we did before. So 20 m will 'need' 8 cm and 15 cm will need 6 cm.

Method

1 Draw a line OP 9 cm long due east, and from it cut off OX 8 cm long to represent 20 m. Place an arrow on OX from O to X to show the direction of the vector.

2 At O construct or measure off an angle of 90° below OX. Draw this 90° line longer than the required 6 cm. From it cut off OM 6 cm long to represent the 15 m vector due south. Put an arrow pointing south on this line to show that it represents the vector $\overrightarrow{OM}$.

3 Complete the parallelogram with OX and OM as adjacent sides by using X as centre and drawing an arc below OX of radius 6 cm and another arc of radius 8 cm with M as centre.

4 Where these two arcs intersect will be the fourth corner of the parallelogram (here a rectangle). Denote this point by R and place an arrow directed from O to R to show the resultant vector $\overrightarrow{OR}$.

5 In this parallelogram method $\overrightarrow{OR}$ represents the resultant of the two vectors $\overrightarrow{OX}$ and $\overrightarrow{OM}$. Measure $\overrightarrow{OR}$ and use the scale to calculate its value (see figure 5.6).

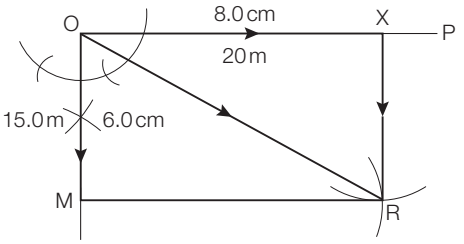


Figure 5.6 Using the parallelogram method to find the resultant of two vectors.

If this construction is done properly the resultant given by the diagonal of the parallelogram should be 10 cm long, resulting in a resultant of 25 N.

Note that whereas the triangle can be used to find the resultant of any two co-planar vectors, whether they are collinear or not, the parallelogram method cannot be used if they are collinear. The difficulty in using the parallelogram method is encountered when you try to find the resultant by drawing the diagonal of a parallelogram which you will not have if the vectors are collinear.

Notice also that the triangle method is shorter in that the amount of construction to be done is less that it is in the parallelogram method.

These are oblique forces.

Worked example 5.1

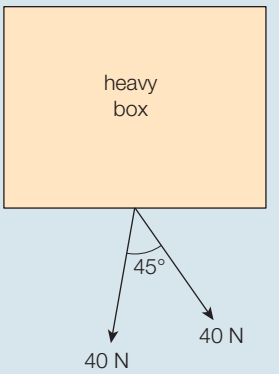


Figure 5.7

A heavy box is pulled along a horizontal floor by two horizontal ropes (figure 5.7). The tension in each rope is 40 N. The ropes are inclined at 45° to each other. Use a scale drawing to find: (i) the total force pulling the box and (ii) the direction of that force.

To obtain the angle of 45° it might be best to construct 90° and then bisect it. Alternatively, you may use a protractor to obtain the angle. It is better to use a ruler and compasses rather than a protractor, however.

Solution

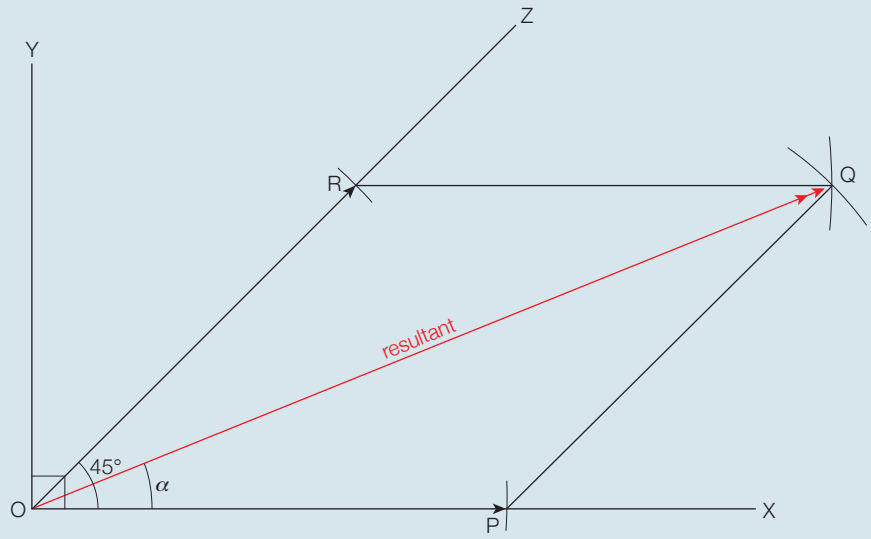


Figure 5.8 Parallelogram method (not to scale).

These are oblique forces.

- Choose a suitable scale. Take 1 cm (of paper) to represent 5 N, because 40 N can be divided by 5 N without leaving a remainder. The 40 N will therefore be represented by 8 cm.
- Construct an angle of 90°, using arms OX and OY of length greater than 8 cm, say 10 cm.
- Bisect the 90° angle, again using a line OZ of about 10 cm as bisector. This gives the desired angle of 45° (figure 5.8).
- On OX and OZ, cut off OP and OR each equal to 8 cm (as shown).
- With P as centre, draw an arc of radius 8 cm, and with R as centre, draw another arc with the same radius (8 cm) in such a way that the two arcs intersect at Q.
- Join OQ.
- Place arrows on the lines OP, OR and OQ, all pointing away from the origin, O. Measure OQ, and use the scale to find the force it represents.
- Measure the angle marked α in the diagram.

The total (or resultant) force is of magnitude 74 N and it makes an angle of 22.5° with each of the forces.

I.T.Q 1

Why should the resultant force $\overrightarrow{OQ}$ in Worked example 5.1 make equal angles with forces $\overrightarrow{OP}$ and $\overrightarrow{OR}$? In what direction will the box move along the floor?

Method

The solution to Worked example 5.1 may be obtained using the triangle method as follows.

1 Decide on a suitable scale (1 cm to 10 N) and draw the first vector, $\overrightarrow{OP}$, say, as you did in the parallelogram method (see figure 5.9).

2 Draw the second vector the same length as the first, starting from the tip or 'tail', P, of the first and making an angle of 45° with $\overrightarrow{OP}$ and above it. Denote this vector by $\overrightarrow{PQ}$.

3 Complete the triangle OPQ. This third side, $\overrightarrow{OQ}$, represents the desired vector.

4 Measure its length and use the scale to find the force it represents.

Don't forget to put the direction arrow on each vector that represents a force!

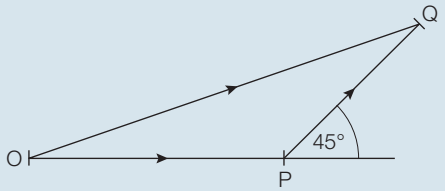


Figure 5.9

To help you to understand this method better, we will do another example.

Worked example 5.2

A girl is lying in a hammock. The hammock is supported by two ropes whose tensions are each 600 N. The ropes make an angle of 120° with each other. Find by drawing and measurement the magnitude and direction of the resultant force supporting the girl and the hammock.

Solution

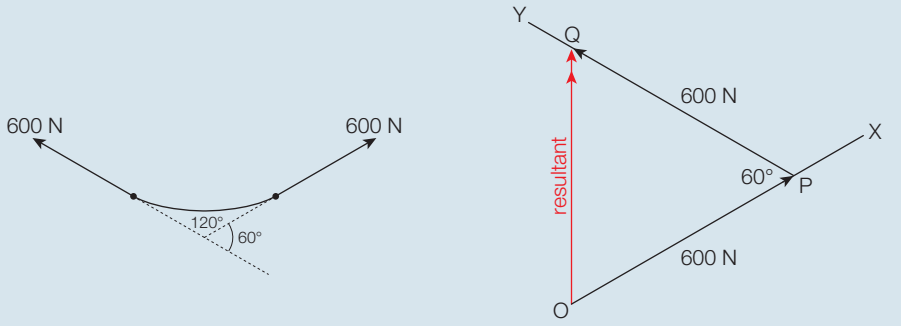


Figure 5.10 Triangle method (not to scale).

First do a quick sketch of the hammock being supported by the two ropes as in figure 5.10. Put in the force directions in the ropes. Now do the following:

- Choose a suitable scale, for example, 1 cm to represent 100 N. The first vector will therefore be 6 cm long.
- draw a line OX roughly parallel to one tension about 8 cm long (longer than 6 cm, the length of the vector to be drawn). Mark off 6 cm. Call this length $\overrightarrow{OP}$ (see figure 5.10).
- Since the vectors have an angle of 120° between them, from P draw a line about 8 cm long making an angle of 60° with this line.
- With centre P, mark off PQ equal to 6 cm.
- Join OQ. This line $\overrightarrow{OQ}$ will represent the resultant of the two vectors.
- Measure the line OQ and use your scale to find the size of the force it represents.

From the diagram it is clear that the angle the resultant makes with the two ropes will be the same for both. This angle should be 60°.

Combining parallel vectors

parallel vectors ›
anti-parallel vectors ›

We met the case of **parallel vectors** in the diagram of figure 5.1 (a) and that of **anti-parallel vectors** in figure 5.1 (b). In these cases the vectors were along the same line. If, instead of two displacements, we had two forces acting on a body along different lines (or different lines of action), the resultant would be the same as if they acted along the same line.

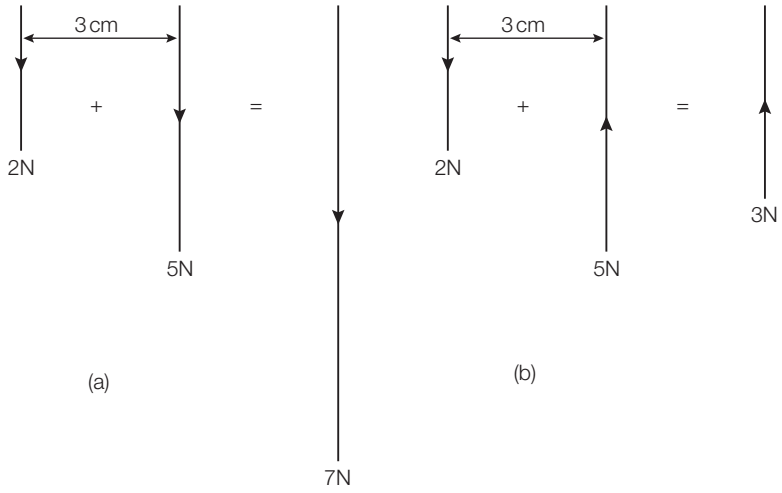


Figure 5.11

So a force of 2 N acting 3 cm from a parallel force of 5 N (figure 5.11 (a)) would have a resultant of 7 N in the direction of the forces, whereas the same force of 2 N acting the same 3 cm from an anti-parallel force of 5 N (see figure 5.11 (b)). would have an overall resultant of 3 N in the direction of the predominant force of 5 N.

Adding these vectors graphically is done by drawing the first vector (using an appropriate scale) in a stated direction. Then from the end of the vector, draw the second. If the vectors are parallel, then both go in the same direction. If they are anti-parallel, then one goes in the opposite direction to the other.

ITQ2

Calculate the resultant of the pairs of vectors shown below. The first vector in each pair acts from left to right across the paper. State the direction of the resultant in each case.

(i) 30 N and 50 N (parallel).

(ii) 4 mm s^{-1} and 7 mm s^{1} (anti-parallel).

(iii) 1 m s^{-2} and 1 m s^{2} (anti-parallel).

Worked example 5.3

Two forces of 3 N and 5 N respectively act at the corners A and B of a square lamina ABCD as shown in figure 5.12. The 3 N force acts in the direction DA and the 5 N force acts in the direction CB. A third force of 6 N acts at the corner C in the direction DC. See figure 5.12 (a).

Calculate:

(a) the resultant of the forces acting at A and B,

(b) the resultant of the three forces acting on the lamina in magnitude and direction.

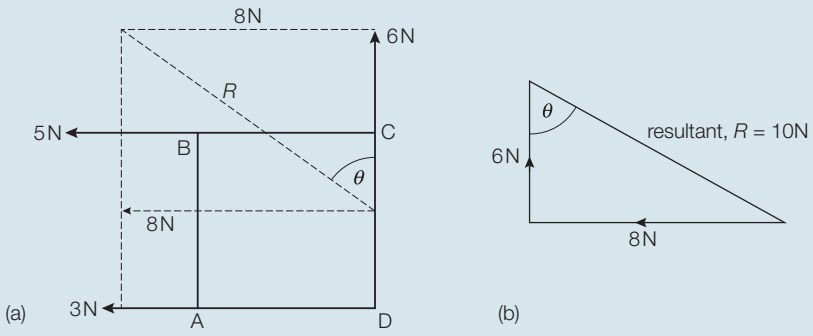


Figure 5.12

Solution

(a) Since the 3 N and 5 N forces are parallel, their resultant = 8 N.

(b) This resultant is at right angles to the 6 N force, and so their resultant, R, is given by $R^2 = 8^2 + 6^2$

$= 64 + 36$

$= 100$

and so $R = 10$ N

From the diagram of figure 5.12 (b):

$\tan\theta = \frac{8\text{ N}}{6\text{ N}}$

$= \frac{4}{3}$

$= \tan 53.1°$

and $\theta = 53.1°$

So the overall resultant of the forces is a force of 10 N acting at an angle of 53.1° with DC.

Resolving a vector into two components

CHAPTER 7

To resolve a vector along two directions is to find two values of that vector which, when made to act along those directions, have the same effect as the one original vector. Most often, for convenience, we resolve a vector into two directions at right angles to each other. Resolving vectors is extremely useful and sometimes necessary when studying the equilibrium of bodies (chapter 7). There are two methods that could be used:

- the rectangle method;
- calculation.

Use the method that you find easier, unless you are instructed otherwise!

The rectangle method

A vector representing the force to be resolved is drawn as the diagonal of a rectangle. The diagonal makes any desired angle with one of the sides of the rectangle. The sides of the rectangle built on that diagonal represent the components of the vector.

We will use an example to show the method in detail.

Worked example 5.4

Resolve a velocity (not speed!) of 20 m s^{-1} into two perpendicular directions so that one component makes an angle of 30° with the 20 m s^{-1} velocity.

Solution

- Choose a suitable scale. We will use 1 cm to represent 2 m s^{-1}. This will give a large diagram, which will increase accuracy. So 10 cm will represent 20 m s^{-1}.
- Construct an angle of 90° with arms longer than 10 cm. We will use 11 cm. Call the arms OX and OY (figure 5.13). These are the directions along which the resolved velocities will act.

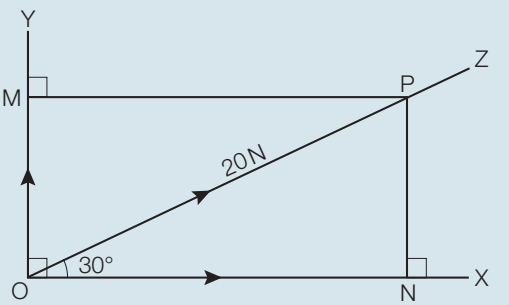


Figure 5.13 Rectangle method (not to scale).

MATHEMATICS: geometry

- At O construct an angle of 30° (within the right angle) with OX. Call this line OZ. The constructed arm OZ should be longer than 10 cm.
- Now along OZ mark off OP = 10 cm. This line $\overrightarrow{OP}$ represents the given velocity of 20 m s^{-1}.
- From P construct perpendiculars to OX and OY to meet these lines at N and M, respectively.
- Measure ON and OM. These represent the two components of the given velocity.
- Use the scale to work out the size of the components.

Now carry out the construction for yourself and see whether you get answers that are close to 17 m s^{-1} and 10 m s^{-1} for $\overrightarrow{ON}$ and $\overrightarrow{OM}$ respectively.

General rule for components

The resolved part of a vector F in a direction that makes an angle θ with the vector F is $F\cos\theta$. The other component of the force is $F\sin\theta$.

Using calculation

We do not always need to use graphical methods. Sometimes a little trigonometry will do the job. Look at figure 5.14, which is based on the 'vector rectangle' taken from figure 5.13. From the diagram,

$$\frac{ON}{OP} = \cos 30°$$

and so

$$ON = OP \cos 30°$$

Similarly,

$$OM = OP \cos 60°$$

or

$$OM = OP \sin 30°,$$

since $\cos\theta = \sin(90^0 - \theta)$

So $ON = 20 \cos 30° = 17.3\,N$

and $OM = 20 \sin 30° = 10.0\,N$

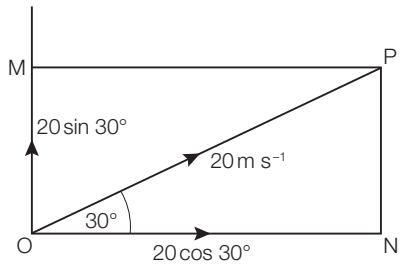


Figure 5.14

MATHEMATICS: trigonometry

Chapter summary

- Two vectors meeting at a point (and therefore in the same plane) may be added (or combined) to give a resultant.
- A resultant of a set of vectors is that vector which has the same effect on a system as the set of vectors acting together.
- Two co-planar vectors may be combined by using:
  - the rectangle method;
  - the triangle method.
- A single vector may be replaced by two separate vectors at right angles to each other.
- These vectors are called the components or resolved parts of the single vector.
- There are two ways of resolving a vector:
  - using a rectangle;
  - using calculation.

Answers to ITQs

ITQ1 Since the forces are equal, there is no reason why the resultant should be nearer to one than to the other. In other words, there should be symmetry about the resultant. The resultant should therefore be half-way between the two forces. The box will move along the line of action of the resultant.

ITQ2 **(i)** 80 N to the right; **(ii)** 3 N to the left; **(iii)** 0 N

- The resolved parts of a vector, P, may be obtained as follows:
  If one component is to be inclined at an angle θ to the vector P, the components will be $P\cos\theta$ and $P\sin\theta$.

Examination-style questions

1 A circular track has a radius of 120 m. An athlete runs half-way round the track and stops.

(i) How much distance has she covered?

(ii) Calculate her displacement.

She takes 54 s to cover the distance.

(iii) Calculate her average speed, using the formula

$$\text{average speed} = \frac{\text{distance covered}}{\text{time taken}}$$

(iv) Calculate her average velocity using the formula

$$\text{average velocity} = \frac{\text{displacement}}{\text{time taken}}$$

2 By drawing and measurement, resolve the vector P into two perpendicular components, one of these making an angle θ with P, where

(i) $P = 20$ N, and $\theta = 20°$

(ii) $P = 400$ km h^{-1} and $\theta = 60°$.

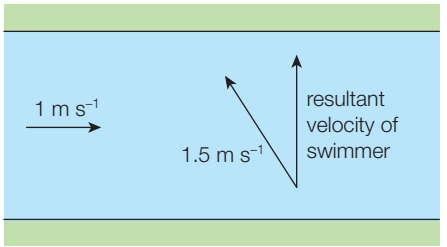
Use both the parallelogram method and the triangle method for each construction and compare your results.

(iii) Which do you consider the more reliable method? Why?

3 Two tugs pull a ship by means of two horizontal cables, each with a tension of 10 000 N. The angle between the cables is 60°. Find, by drawing and measurement, the total force pulling the ship.

4 A swimmer can swim at 1.5 m s^{-1} in still water. He wishes to cross a river in which there is a current flowing at an average speed of 1 m s^{-1}.

Find, by drawing and measurement, the direction in which the swimmer should swim in order to cross the stream at right angles. (Hint: use the fact that the resultant velocity of the stream and the swimmer should be at right angles to the bank of the stream.)



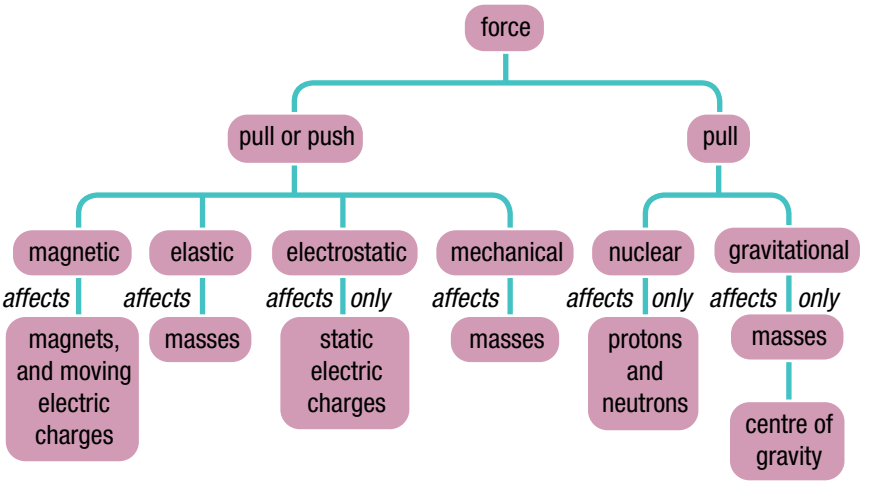
5 An aircraft which can travel at 500 km h^{-1} in still air points its nose due east and flies in a wind moving in a north easterly direction at 50 km h^{-1}.

By means of a scale drawing, find the resultant velocity of the aircraft and the angle its path will make with due north.

6 Forces

By the end of this chapter you should be able to:

- appreciate that a force is always either a push or a pull
- recall some ways in which forces are produced
- understand that a force may change the size, shape or motion of a body
- understand that forces generally act in pairs, called 'paired forces'
- recall that paired forces act on two different objects of a system at the same time
- understand the concept of a derived quantity and how such a quantity is obtained
- recall that some derived quantities are given special names, generally after famous scientists
- understand the meaning of the term 'field of force' as applied to gravitational forces
- define the term 'gravitational field strength'
- use the term 'gravitational field strength' to calculate the weight of an object
- distinguish between the 'mass' and 'weight' of a body
- understand the meaning of the term 'centre of gravity' of a body
- understand why 'centre of gravity' may also be called 'centre of mass'
- understand the significance of the centre of mass (or the centre of gravity) of a body in solving problems



How various forces are produced

Why does your pen move along the paper as your write? Because you push it along the paper. If you do not push it, the pen will not move. Why does a bow move and change shape when you pull the string before releasing an arrow? Because you exert a pull on the string (and also a push on the bow). If you do not pull the string, the bow and the string will not move and change their

shape. Why does a cricket ball change direction when you hit it? Because you have pushed it with the bat. So this shows pulling can cause bodies to move and to change their shape, and pushing can cause a body to move or come to rest or change its direction. The push on the pen, the pull on the bowstring and the effect of the bat are all examples of **forces**.

force ›

A force is any influence which changes a body's state of rest or its motion or any influence which causes a body's shape or its dimensions to change.

Forces are produced in a number of quite different ways. Some of these occur in nature; others generally come about as a result of human action. Whatever the influence, the term used to describe a force depends on the way in which the force is brought about. The different types of force we will meet in our study of physics are:

- *mechanical* – the force between surfaces that are in contact with each other;
- *elastic* – the force in materials which can stretch or be compressed to a length different from their natural length;
- *gravitational* – the attractive force between masses, however small these masses may be;
- *electrostatic* – the force between bodies which carry an electric charge;
- *magnetic* – the force between bodies that are magnetised;
- *nuclear* – the force between the neutrons and protons found in the nuclei of atoms.

Contact forces and non-contact forces

Each of these six types of force belongs to one or the other of two main categories namely contact forces and non-contact forces. This classification is shown in the classification map in figure 6.1, which shows the *contact forces* (mechanical forces and elastic forces) and *non-contact forces* (gravitational forces, electrostatic forces, magnetic forces and nuclear forces). Each of these types of force is produced in a particular way.

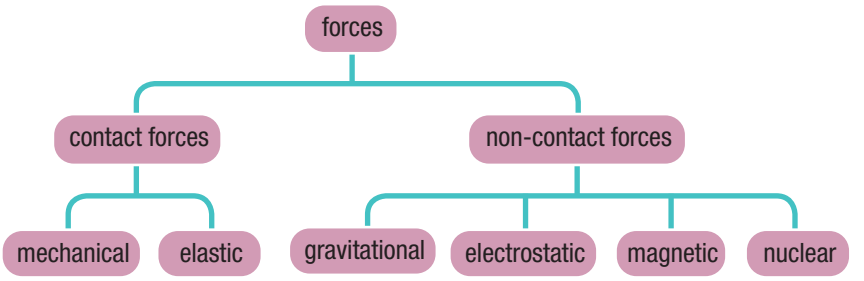


Figure 6.1 Classification map for forces.

Contact force

contact force ›

A **'contact' force** is a force that exists between objects that are in contact. There are countless instances in everyday life of such forces being involved. Such forces are present, for example, when we stand, walk, use a rope to pull on a tree trunk or lean against a surface.

Non-contact force

non-contact force ›

A **'non-contact' force** is a force that exists between two objects that may be separated from each other. We describe this effect as 'action at a distance' since, although the objects are separated from each other, there is still a force between them. Examples of non-contact forces are gravitational, electrostatic and magnetic forces.

Mechanical forces

These forces are produced when two bodies come into contact with each other. The action of one body, A say, pressing against another body, B, produces a thrust (a push) and to this push of A on B there will be a resulting 'push-back' of B on A. These two forces are equal and they act in opposite directions. They are called 'paired forces'. Such forces occur constantly in our everyday experience. It is a very common type of force and it occurs in different forms and in different circumstances. In almost every one of our everyday activities this type of force is present. As long as we touch objects, use instruments and appliances, lie, sit, etc., one form of this type of force or another plays an important part. We use one mechanical force to walk, and meet another when we do 'push-ups' and yet a different type when we swim.

What are the different forms of mechanical force we use in performing our daily routine?

The elastic force

Search the internet to learn more about what strands and bands are and how they were/are used in exercising.

Internet search terms: strands/bands

The other type of contact force is the elastic force. Although it is not as common as the first, it is nevertheless very familiar. (Just imagine life without rubber bands!) What did we use instead? This is the type of force upon which acrobats and gymnasts depend when they use the trampoline. It is the type with which the exercise enthusiasts were concerned when they 'pulled strands' (not so common nowadays) or used resistance bands.

What are some of the situations in which the elastic force is present?

Gravitational force

This force is with us all the time. Although we may not be aware of it, it is this force that keeps us on the Earth's surface or on any surface on which we happen to be standing. Without this force it would be impossible to walk or to run or to stand. Without this force we would probably float about in the air like astronauts in their capsules in space. Have you seen a video of Neil Armstrong and Edwin Aldrin, the American astronauts, 'walking' on the Moon? Were they really walking as we know it? If not, find out why.

The types of force with which we are most familiar are probably the first three listed above namely mechanical, elastic and gravitational. We have already discussed the first two, both of which are contact forces. We go now to the non-contact forces.

What is the importance of the gravitational force in our everyday lives?

The last three types of force, electrostatic, magnetic and nuclear, are much less familiar than the first three. Although we may not be aware of their existence or their importance, they nevertheless play a very important part in all the technology we use today. Much of it depends on the important part played by electrons and protons in the electrical and electronic components used in present-day technology.

As a group, investigate how electrons allow your television set and mobile devices to work.

paired forces ›

All of the six types of force mentioned above have a common feature – they occur in **pairs**. Whenever one type of force acts in one direction on one part of a system, it will be accompanied by another force of the same type acting in the opposite direction on another part of the system.

Internet search term: Moon landing

Line of action of a force

The line along which a force is acting is called 'the line of action' of the force.

A force (a vector quantity) is normally represented by a straight line with an arrow on the line pointing in the direction of the force (→). In other words, the arrow representing the force vector points along the line of action of the force.

We now consider each of these three types of non-contact forces – gravitational, electrostatic and magnetic – in more detail.

Gravitational force

Mass and weight

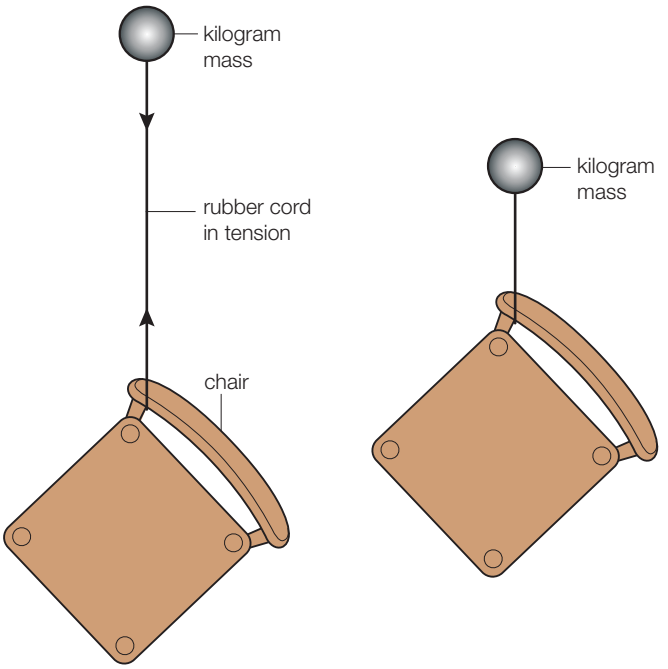
Since gravitational force is the type of force that exists between masses only, we need first to understand the nature of 'mass' or what 'mass' means. Some say that mass is 'quantity of matter', but we will agree that this 'definition' can be very misleading when we 'consider whether a large block of 'Styrofoam' has more mass than a handful of soil.

What is mass?

Consider the following demonstration to find out which of two objects, a chair and a kilogram mass, will better maintain its position when they are both subjected to the same force.

Materials required

A chair, a 1 kg mass, an elastic cord (such as a strip of bicycle tubing about 30 cm long) and a 'smooth' laboratory bench.



(a) Before release (viewed from above)

(b) After release

Figure 6.2

Method

1. Place both chair and kilogram mass on the smooth bench top.
2. Securely attach one leg of the chair to one end of the elastic cord and the kilogram mass to the other. Ensure that the elastic cord is so attached that when stretched it would be parallel to the bench top.
3. Pull the chair and the kilogram mass apart to a separation of between two or three times the unstretched length of the cord. Hold them apart at this separation (figure 6.2 (a)).
4. Now release both chair and kilogram mass at the same time (figure 6.2 (b)).

Results

- What did you notice happening to the block and the kilogram mass when they were released?
- Did they both remain at rest or did they both move?
- If they both moved, which one, the chair or the kilogram mass, moved through the smaller distance?

Inference

The inference from the observation made in this case is that the body that moved the smaller distance was better able to maintain its state of rest. It is therefore said to possess more **inertia** and, since **mass** of a body is taken to be related to the ability of that body to 'resist' a force and maintain its position, we must conclude that the chair has more inertia than the kilogram mass since it moved a smaller distance that the kilogram mass, both being subjected to elastic forces of the same size.

An inference is a first or (intermediate) conclusion formed on the basis of observation and evidence. This inference often leads to a further final conclusion which is always related to the aim of an experiment or investigation.

inertia mass ›

Conclusion

The chair is heavier than the kilogram mass.

Discussion points arising out of the demonstration

Since the two masses were being compared, all variables which control the behaviour of the masses (e.g. the pull of the elastic cord, i.e. the tension), and friction from the bench top must be the same for both bodies (controlling variables!) Were they? How did the experiment attempt to achieve this?

If you think that the force that pulled the bodies was not the same for both, state why you think that one could have been be greater. Which of the two bodies would experience the greater force? Why?

You will need your answers to these questions to help answer ITQ8.

The mass of a body is a measure of the inertia of that body, i.e. a measure of the ability of that body to maintain its position under the action of a force. The mass depends on the number and size of the atoms in the body. It will not change if the body is taken from one place to another. The only way in which the mass of a body will change is if some of the matter of which it is made is removed.

The magnitude of a gravitational force

The magnitude of the gravitational force between two bodies increases as the mass of each of the bodies increases, but decreases rapidly as their separation increases. To understand how this forces varies with the size of the masses and their distance apart, see figure 6.3.

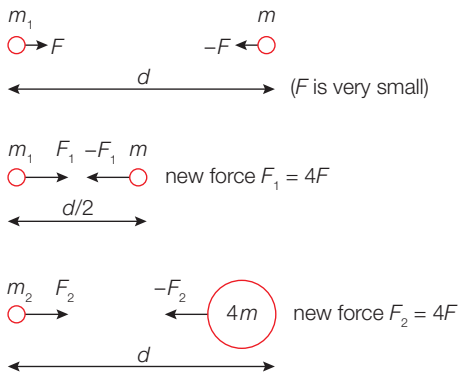


Figure 6.3 (a) When separation is *d*, gravitational force = *F*. When separation changes to *d*/2, force increases to 4*F*. Force varies inversely as the square of separation. (b) When at separation *d* one mass is increased by 4, the gravitational force increases to 4*F*. Force varies directly as mass size.

ITQ1

The force of attraction between two small brass spheres hanging close together from strings of equal length is so very small, that it can be ignored. What is the evidence that the gravitational force between the two spheres is negligibly small?

ITQ2

When one of the strings in ITQ1 is cut, the sphere that was connected to it falls to the ground. What does this observation tell us? Explain why the sphere falls.

Gravitational forces are generally so very small that we may well be unaware that they exist at all. This is so only because the masses we meet ordinarily are very small compared with the mass of the Earth, which exerts an appreciable gravitational force even on objects of small mass (for example a grain of sugar!)

The gravitational force between two objects acts on both of them (figure 6.4). If the objects in question are A and B, then A attracts B with a certain force, *F* in one direction, and B attracts A in the opposite direction with a force of the same size (and so –*F*). This is what we mean when we say that forces occur in pairs. The presence of one force, *F*, gives rise to the pairing force, –*F*. Note that these two forces act on different bodies and are of the same nature namely, gravitational. Force *F* acts on body B and force –*F* acts on body A.

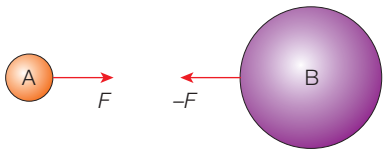


Figure 6.4

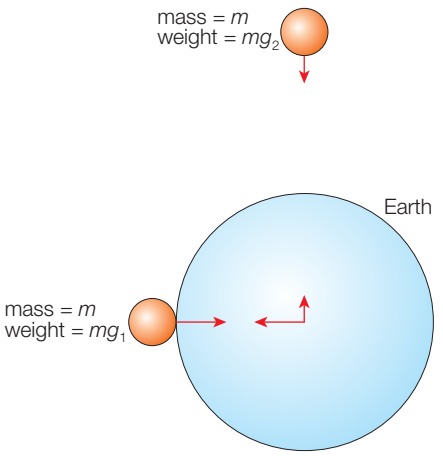


Figure 6.5 Above the Earth's surface, because $g_2 < g_1$, the weight of the mass m is less than it is at the Earth's surface.

gravitational field ›

gravitational field strength ›

g ›

Mass and gravitational force

The Earth, with its very large mass, must have a very large inertia. Objects on Earth, on the other hand, all have a small inertia by comparison, because of their relatively small mass. The force between the Earth and another body will act both ways, one on the body (due to the Earth's pull) towards the Earth and the pairing force on the Earth (due to the body's pull) towards the body (see figure 6.5). Since the inertia of the earth is so very great, its response to the pull of bodies on Earth will be negligible and can be ignored. However, the force the Earth exerts on the bodies on the Earth will be large enough to cause them to move appreciably. Their movement will be in the direction of the attractive force, which is towards the centre of the Earth. All bodies released from a point above the earth will therefore fall towards the centre of the earth. This direction of fall is what we know as the vertical.

The space around the Earth in which masses feel a pull from the Earth is called the **gravitational field** of the Earth.

The force which the Earth exerts on one kilogram of mass is called the **gravitational field strength** of the Earth. This quantity is represented by the letter g.

The S.I. unit of g and the S.I. unit of weight

The pull (gravitational force) of the Earth on 1 kg of mass is g. The S.I. unit of g is therefore

$$\frac{\text{S.I. unit of force (N)}}{\text{S.I. unit of mass (kg)}} = \text{newton per kilogram (N kg}^{-1}\text{)}$$

weight ›

The force an object experiences due to the Earth's pull is called its **weight** (symbol W). It is important to avoid the mistake (common among students) of referring to weight of a body as '*gravity*'. Gravity is the effect which gives rise to weight. Note also that 'g' does not stand for 'gravity'.

Since weight is a force, then its S.I. unit is the newton.

The value of g

Weight, $W = mg$

If 1 kg is taken as having a weight of 10 newtons, or g, then a mass of m kg will have a mass of mg newtons.

On the surface of the Earth, the value of g can be taken to be 10 N kg^{-1}. So a mass of 1 kg is assumed to experience a gravitational force (its weight) of 10 N. Away from the Earth's surface the value of g becomes smaller with height above sea-level (figure 6.5).

ITQ3

The value of g at the surface of the Moon is 1/6 of the value at the Earth's surface. Taking the latter to be 10 N kg^{-1}, calculate the weight of an astronaut of mass 72 kg on (i) the Earth's surface, and (ii) the Moon's surface.

Earth's gravitational force, centre of gravity and centre of mass

centre of gravity (C.G.) ›

Centre of gravity is often abbreviated as C.G. in books.

Bodies are made up of particles of matter. Each of these particles is pulled to the Earth by a gravitational force. The resulting total gravitational force pulling the body is called the weight of the body.

It is very convenient in doing calculations involving the weight of a body to regard its weight as acting from one point only (figure 6.6). This point at which the total weight of the body is assumed to act is called the **centre of gravity** of the body.

centre of mass (C.M.) ›

Centre of mass is sometimes abbreviated as C.M. in books.

Since the gravitational force acts only on masses, we could call the centre of gravity of a single object by another name, the **centre of mass**. This implies that the centre of mass is that point where the entire mass of the body may be taken to be concentrated.

ITQ4

The gravitational field strength on the earth's surface, g, is taken as 10 N kg^{-1}, and on the moon's surface as $g/6$. What mass will have a weight of 6 N on the moon?

We see, then, that the total gravitational force on a body acts from the centre of mass of the body. In calculations involving weight, this force is always taken to act from the centre of mass of the body. Note that the centre of mass of a body need not be within the solid part of the body (figure 6.7). (For example, the centre of gravity of a tennis ball is at the centre of the space inside the ball.)

If a body has axes of symmetry and is of uniform density, its centre of gravity may be found by using its symmetrical features. Thus the centre of mass of a rectangular box is the point where the body diagonals meet or where any two axes of symmetry meet.

If a body does not have axes of symmetry and is not of uniform density, its centre of gravity may be found by experiment. See an account of the experiment described in chapter 7.

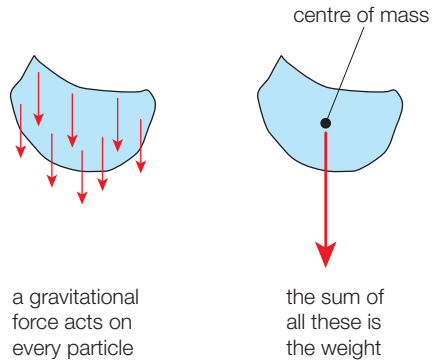


Figure 6.6

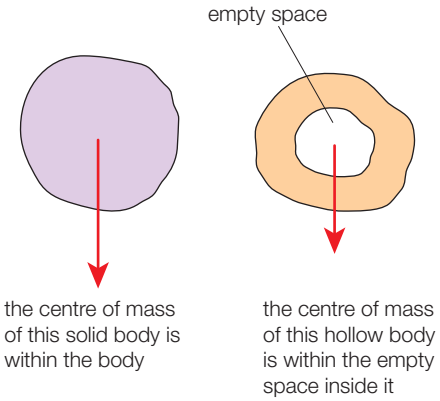


Figure 6.7

CHAPTER 7

Mechanical forces

mechanical force ›

Mechanical forces are forces that exist between objects in contact. Such forces may be of quite different natures. Examples of mechanical forces are:

thrust ›

- **thrust** – a 'push' exerted on a surface (see figures 6.8 and 6.9);

normal reaction ›

- **normal reaction** – the thrust (called a normal reaction), resulting from an initial thrust between two solid surfaces in contact, the thrust being always perpendicular (or normal) to the surface exerting it (see figures 6.8 and 6.9);

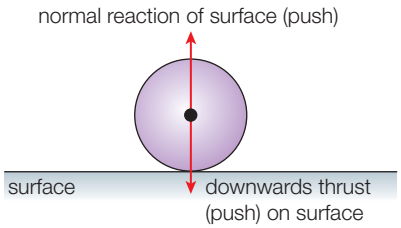


Figure 6.8

tension ›

- **tension** – a force in a thread, string, or chain, called tension, which can act in opposite directions (on different bodies) (see figure 6.10);

Archimedean upthrust ›

- **Archimedean upthrust** – (after Archimedes, a Greek philosopher) an upward thrust exerted by fluids (liquids and gases) on bodies immersed in them (see figure 6.11);

Internet search term: the story of 'Eureka!'

friction ›

- **friction** – a force between and parallel to two solid surfaces which are in contact and are moving relative to each other (see figure 6.12). This force always acts in the opposite direction to that in which motion is occurring;

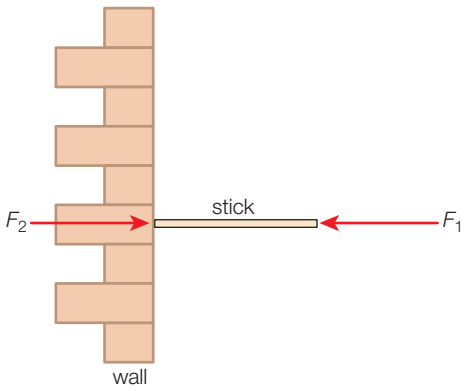


Figure 6.9

drag ›

- **drag** – a force between a solid surface and a fluid called 'drag' (see figure 6.13). This force, like friction between solid surfaces, also acts to oppose relative motion.

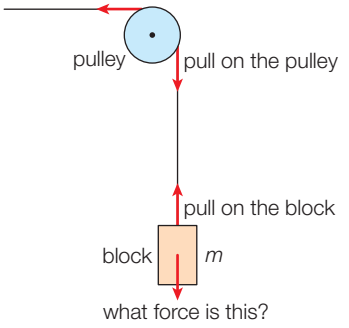


Figure 6.10

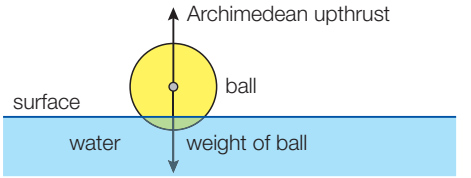


Figure 6.11

ITQ5

Two men are carrying a pole, each on his left shoulder. Choose the correct name for the forces described below:

(i) The force acting downwards on each man's shoulder. (*weight, normal reaction, friction, thrust*)

(ii) The force acting upwards on the pole at each man's shoulder. (*weight, normal reaction, friction, thrust*)

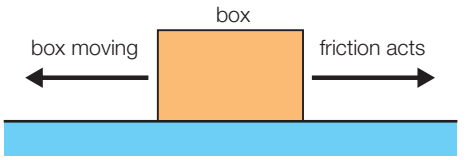


Figure 6.12 Friction acts to the right between surfaces in contact.

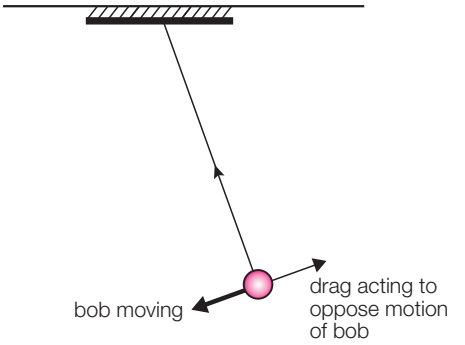


Figure 6.13

Non-contact forces

electrostatic force ›

An **electrostatic force** may be produced by rubbing two non-conducting materials together.

Practical activity 6.1

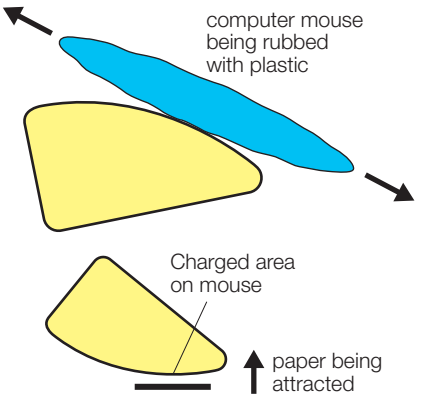


Figure 6.14 After being rubbed with the piece of plastic the mouse attracts a piece of paper.

Producing an electrostatic force

In this activity you will produce an electrostatic force by rubbing two materials together.

You will need:

- computer mouse
- small piece of plastic
- a very tiny piece of paper (about 2 mm^2).

Method

1 Rub the top of the computer mouse vigorously with a piece of plastic (figure 6.14).

2 Hold the part of the mouse that was rubbed about 1 cm above the paper. (Note that both the paper and the material of the computer mouse (plastic) are electrically non-conducting.)

Observation

The strip of paper is picked up by the mouse and sticks to it.

Inference

The mouse exerted an attractive force on the paper.

Conclusion

This force was not a contact force, for there was no contact between the mouse and the paper when the latter was picked up. The force was therefore gravitational, or electrical or magnetic. How would you show that it must have been electrical and not magnetic or gravitational? It was clearly a non-contact force.

magnetic force

Magnetic forces

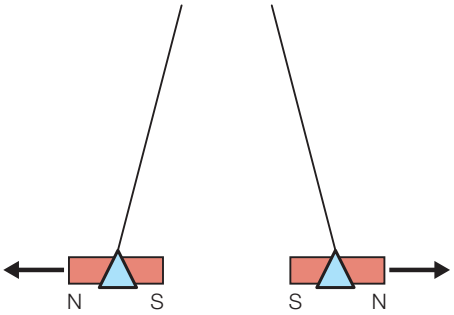


Figure 6.15 These two magnets are repelling each other with a magnetic force.

ITQ6

How could you convince someone that the force acting in Figure 6.15 is not, possibly, also electrical or gravitational in nature?

Electromagnetic force

An electromagnetic force is a magnetic force produced by an electric current rather than by a magnet.

You have probably seen the arrangement shown in figure 6.15. This shows two bar magnets hanging from two strings in line with each other with the north poles facing away. These magnets are repelling (pushing away) each other. The force of repulsion is a magnetic force, which can act at a distance. This is clearly a non-contact force.

In figure 6.16, a straight wire is passed at right angles through a horizontal sheet of stiff cardboard and a small plotting compass is placed on the cardboard with its needle pointing away from the wire. When a strong current is switched on in the wire, the needle is seen to turn away from its previous direction. Since the only type of force that will affect the compass needle is a magnetic force, we conclude that the current in the wire exerts a magnetic force on the needle of the plotting compass. The current-carrying wire behaves like a magnet. It produces a force that can cause movement in another magnetic object, such as the compass needle.

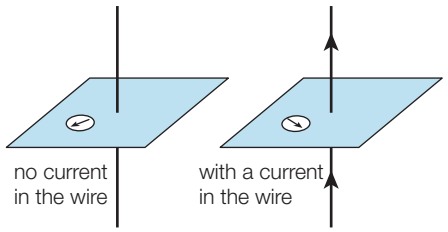


Figure 6.16 A current-carrying wire behaves like a magnet.

This force is an example of an electromagnetic force. It is a non-contact force.

Elastic forces

If an elastic band is pulled between the first finger and thumb of each hand, it will stretch. The pull exerted by each hand changes the length of the band. As the hand pulls one end of the band in one direction, the tension in the band exerts an opposite force on that hand. This is another example of forces acting in pairs and on different bodies. In the same manner, a hanging rubber cord will stretch if a ball is attached to it, becoming longer and thinner (figure 6.17).

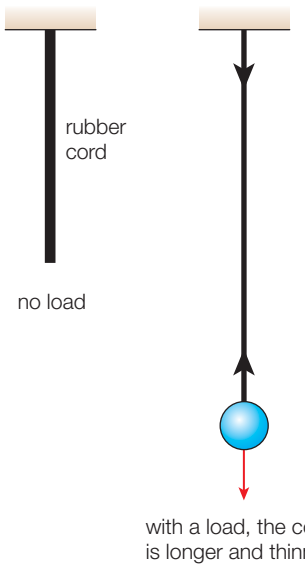


Figure 6.17 An elastic force can change the dimensions of an object.

tensile force
tension

Since this force is applied along the length of the rubber band, we call it a **tensile force** and we say that the rubber band is in **tension**. When the force is removed, the ends of the band move towards each other and the band shortens again. Clearly the force in the band produces a change of dimensions in the band. We call this force an **elastic force**, since it is associated with stretching or elasticity.

elastic force

Other objects that show similar behaviour are springs and thin wires. This is an important type of force, as it is the only one that causes a change of length in the body to which it is applied. We met this force in chapter 3 when we considered the extension of a spring in Practical activity 3.1.

CHAPTER 3

ITQ7

A bar magnet is supported in a vertical position by a short length of thread, AB, an elastic band, BC, with its lower end resting on a horizontal table as shown in figure 6.18.
Name and label each type of force acting on the magnet, indicating by an arrow the direction of each force.

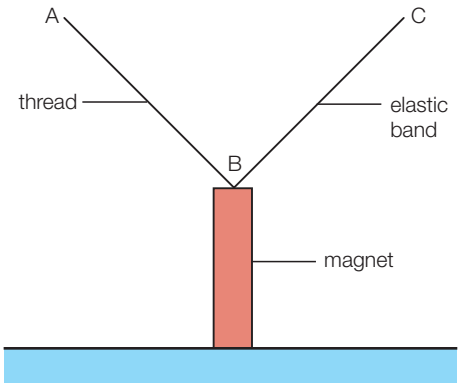


Figure 6.18

ITQ8

Consider again the demonstration discussed on page 79 to demonstrate the meaning of mass. Since the two masses were being compared, all variables which control the behaviour of the masses, for example the pull of the elastic cord (i.e. the tension), and friction from the bench top must have been be the same for both bodies (controlling variables!).

(i) How did the experiment attempt to achieve the same net force and friction for both bodies?

(ii) Were they really the same for both the chair and the kilogram mass? Give a reason for your answer.

(iii) If you think that the net (or resultant) force that pulled the bodies was not the same for both, state which one could have been be greater and why you think so.

Nuclear forces

nuclear force

CHAPTER 33

Nuclear forces act between neutrons and protons within the nuclei of atoms (see chapter 33). They are the forces that hold the protons and neutrons of the atoms together within the nucleus (figure 6.19). These are immensely strong forces and they act only over short distances within the nucleus. You will learn more about this force in your later study of physics. This is perhaps the only type of force that we do not come across in our everyday activities, since it is found only in the nuclei of atoms.

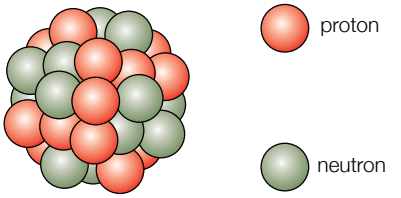


Figure 6.19 Nucleus of an atom.

Chapter summary

- A force is always either a push or a pull.
- Forces act in pairs.
- Forces acting between two bodies always act in opposite directions on these bodies.
- A force produces one or more of the following effects on a body on which it acts:
  - change of position;
  - change of dimensions;
  - change of shape;

  - change of speed in a moving body;
  - change of direction in a moving body.
- There are six main types of force:
  - Gravitational force: the attraction between two masses, whether or not they are separated.
  - Mechanical force: a force that acts between two bodies which are in contact; such a force can be produced in a number of different ways.
  - Electrostatic force: acts between bodies that are electrically charged.
  - Magnetic force: acts between magnetised bodies and current-carrying conductors.
  - Elastic force: occurs in bodies which can be stretched, compressed, or squashed.
  - Nuclear force: binds protons and neutrons together in an atomic nucleus.
- The weight of a body acts from its centre of mass.
- The 'centre of mass' of a single body is the point at which the entire mass of the body can be considered to act. The centre of mass may also be called 'centre of gravity'.

Answers to ITQs

ITQ1 The separation of the strings is the same all along their length.

ITQ2 That there is now a resultant force acting on the sphere when there was none before. Because there is now a resultant force (the weight) causing the sphere to fall.

ITQ3 Weight = $m \times g$. So on Earth, astronaut's weight = $72\,\text{kg} \times 10\,\text{N kg}^{-1} = 720\,\text{N}$ and on the Moon, weight = $\frac{1}{6}$ of $720\,\text{N} = 120\,\text{N}$

ITQ4 $mg = W$, so $m = W/g = 6\,\text{N}/(g/6)\,\text{N kg}^{-1} = 3.6\,\text{kg}$.

ITQ5 **(i)** Thrust; **(ii)** normal reaction

ITQ6 A charged body brought near to the magnet is not attracted or repelled, so force is not electrostatic. A non-magnetic body brought near to the magnet is not attracted, so force is not gravitational. Further, gravitational forces are not repulsive, but attractive.

ITQ7 Your answer should show: **(i)** an arrow along AB directed away from the magnet (a tension); **(ii)** an arrow along BC directed away from the magnet (a tension); **(iii)** an arrow from the centre of the magnet directed vertically downwards (the weight of the magnet); **(iv)** an arrow from the point of contact of the magnet with the table directed vertically upwards (the normal reaction of the table to the magnet).

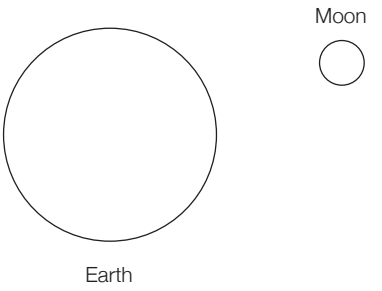
ITQ8 **(i) (a)** By using the pull of the elastic cord, which would have been the same throughout the length of the cord. **(b)** By using a smooth bench top where the friction would be negligible.

(ii) No, since there would have been some friction between the masses and the bench-top.

(iii) Since the friction between the kilogram mass and the bench top would have been the smaller of the two, the net force on the kilogram mass is likely to have been larger.

Examination-style questions

1 The roughly circular path followed by the Moon around the Earth is due to the gravitational force between them.



(i) Copy the diagram and draw arrows showing the force acting:

(a) on the Earth;

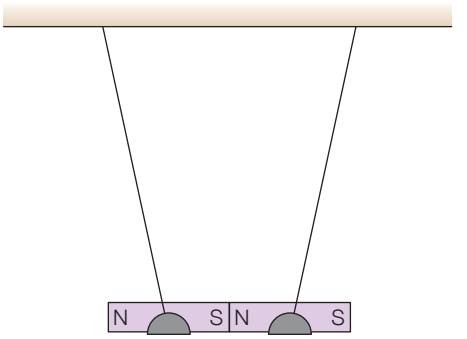
(b) on the Moon.

(ii) What is the relationship between these forces? Imagine a satellite half-way between the centres of the Moon and the Earth.

(iii) Draw arrows showing the forces acting on the satellite due to the mass of the Moon and that of the Earth.

(iv) Which of the two masses will exert the greater force on the satellite? Give a reason for your answer.

2 Two identical bar magnets are suspended in stirrups from two strings so that they are initially in line, with an N-pole and an S-pole facing each other. When released, they remain at rest as shown below.



Copy the diagram and add arrows showing the forces acting on each of the magnets. State clearly which of the forces is/are mechanical, magnetic and/or gravitational.

3 (i) Explain the relationship between the mass of a body and its weight.

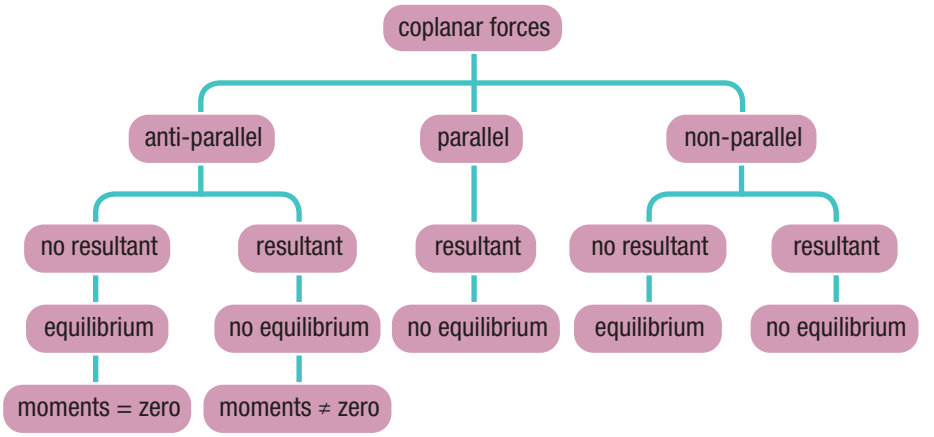
(ii) What do you understand by the statement: 'the value of g on the Moon is 1/6 of that on the Earth'?

(iii) If g on the Earth is 10 N kg^{-1}, what mass (in kg) will weigh 10 N on the Moon?

7 Moments and Levers

By the end of this chapter you should be able to:

- state the meaning of the moment of a force about a point and define it
- use the concept of moments to solve problems
- use the concept of moments to explain the action of common tools and devices
- use the principle of moments to determine the position of the centre of gravity (or the centre of mass) of a body in the shape of a lamina
- use the concept of centre of gravity or centre of mass of a body to examine whether or not that body will be stable
- understand and explain why certain conditions affect the stability of a body in equilibrium



Coplanar forces

coplanar forces

parallel forces

anti-parallel forces

non-parallel forces

In this chapter we are going to discuss the effect on a body of forces which are all in the same plane. Such forces are termed '**coplanar forces**' (figure 7.1). Coplanar forces acting on a body may be put into one of three groups, namely:

- **parallel forces**: all the forces acting in the same direction;
- **anti-parallel forces**: some of the forces acting in one direction and the others acting in the opposite direction – all the lines of action are parallel;
- **non-parallel forces**: the lines of action are not all parallel to one another.

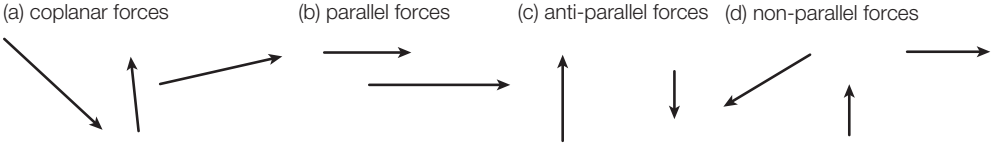


Figure 7.1

CHAPTER 6

We have already learnt the meaning of the word 'resultant' in the context of vectors. We saw that the resultant of a set of forces is that single force which has the same effect as the group of forces acting together. Since we are

about to study the equilibrium of bodies (or the equilibrium of forces as we sometimes say, we must make clear the meanings of terms used in the context of equilibrium:

equilibrium ›

- **Equilibrium** is the state in which a body is not changing in position, composition, shape or movement. Forces which act on a body to keep that body at rest or in equilibrium are said themselves to be at **rest**; the body is in equilibrium under the action of these forces. The size and direction of the forces that are acting will determine whether or not there will be equilibrium.

forces at rest ›

centre of mass ›

- **Centre of mass**. This, as we saw in the previous chapter, is that point where we may consider the entire mass of the body to be concentrated.

centre of gravity ›

- **Centre of gravity**. This is the point at which the pull of gravity acts on the mass of the body.

resultant force ›

- **Resultant force**. The resultant of a set of forces is that one force which has the same effect as the group of forces acting together.



Equilibrium



Figure 7.2

How bodies behave under the action of forces depends not only on the nature of the forces, but also on how the forces are arranged and how large they are. How a sailboat behaves at sea depends on the force and direction of the wind and the waves. How a cricket ball moves when struck by a batsman depends on the stroke played and how hard the ball is hit. It is not only movement that is affected by forces. A hummingbird can remain in one place to collect pollen from a flower (figure 7.2) although its weight continually pulls it towards the Earth. The arrow that an archer is aiming at his target is stationary until he releases it, although the bent bow-string was always forcing the arrow forward.

We first consider whether a body can be at rest (or in equilibrium) under the action of parallel forces. The concept map says 'No'. The reason there can be no equilibrium of parallel forces is that parallel forces (like those shown in figure 7.1 (b)) will always have a resultant, which means that there will be a net force which will produce motion in the body on which the forces act. There can, however, be equilibrium in a set of anti-parallel forces like those in figure 7.1 (c). Perhaps the simplest set of anti-parallel forces is provided by a spring balance supporting a mass. Let us consider the 'statics' of this arrangement.



Suggest three other examples of an object in equilibrium, one from the home, one from the surroundings and one from a form of sport.

In figure 7.3 the lump of plasticine is supported at the end of a spring. Since the plasticine is at rest (in equilibrium), there is no resultant force acting on it and this means that the upward pull of the spring on the plasticine, which is the tension in the spring, is exactly balanced by the downward pull of gravity on the plasticine. So here we have a simple set of forces acting in opposite directions keeping a body at rest. We can use an extension of this arrangement to study the way in which the extension of the spring varies with the tension in the spring, since the tension will be equal to the weight of the object at the end of the spring. This is the subject of the activity described in figure 3.9 of chapter 3, where we tested Hooke's Law. Since the graph of extension of the spring against load was a straight line, we can presume that the extension of the spring, x, varies directly with the tension, T, in the spring, or

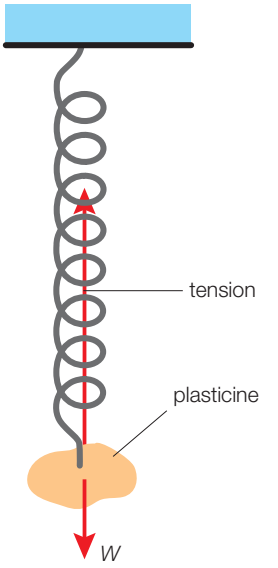


Figure 7.3

CHAPTER 3



$$x \propto T$$

and so $x = kT$ (where k is the proportionality constant)

giving $T = \left(\frac{1}{k}\right)x$

The ratio (any value of T)/(the corresponding value of x) is called the spring constant of the spring being used. It represents the force required to stretch the spring by 1 metre. Since T/x is the inverse of the slope of the graph, we may find the spring constant of the spring used in the activity by taking the inverse of the slope of the graph of x against T. Every spring will have its own value of spring constant, though one may well find that a set of springs belonging to the same batch may have roughly the same spring constant, but not quite the same value.

There is one further point worth mentioning. If you look at the graph of figure 3.10 (see page 39) you will notice that it is only a straight line for loads up to a certain value. If loads are used beyond this value the graph suggests that the spring gets easier to stretch, since you are now getting the same extensions for smaller and smaller additional loads. Beyond the point marked X on the graph Hooke's law is no longer obeyed.

If you remove the load from the spring completely at the point Y, you will find that the spring does not return to its former natural length. The spring has been 'overloaded' – it can no longer behave as it did before. It would not now give the same extension for any fixed increase in load.

Equilibrium of a beam balance

A set of coplanar parallel forces will always have a resultant (which is the sum of the forces) (see figure 5.11 (a)). We could, however, have equilibrium of a set of anti-parallel forces as in figure 5.11 (b) if the two anti-parallel forces were of the same magnitude. It follows then that when the beam of a beam balance remains at rest (the beam as a whole not moving and horizontal), the condition of being balanced is not due to the action of parallel forces at all, but, rather, to the action of the parallel forces acting downwards together with one force acting upwards – three forces, two downward and one upward. These together constitute anti-parallel forces. So where is the third (upward) force? What is it its nature?

The concept map suggests that there can be equilibrium under the action of anti-parallel forces (forces acting in opposite directions). We must therefore look for an upward force acting on the beam and such a force we find at the point of support of the beam – at the pivot or fulcrum. Here the pivot must be exerting an upward force on the beam in order to obtain equilibrium. We can already see, therefore, that when the beam is balanced, the reaction at the pivot must balance the total downward force (weight), or

total upward force = total downward force

This is one condition for a lever to be in equilibrium. This is a requirement if the centre of mass of the balance beam is not moving.

moment of a force ›

If you are familiar with beam balances you will know that there are scale pans hanging at the ends of the beam.

There is a picture of a beam balance in chapter 1 (figure 1.2).

The object to be weighed is placed in the right-hand scale-pan and standard masses (commonly called 'weights') are placed in the other. We then look for balance or equilibrium of the beam in which the balance remains horizontal. How does this come about? We are reminded of the fact that a force can cause an object to move if suitably applied and so we find that the weight of the object on the right (a downward force) has a turning effect on the balance beam which, if there is nothing in the left-hand pan, causes the right hand half of the beam to tip down and the left-hand half to tip upwards. This turning effect of the object is called the 'moment of the weight of the object' and it is referred

to as a 'clockwise moment' because its effect is to turn the beam in a clockwise sense about the pivot. In a similar way, if there is no object in the right-hand pan, but a so-called 'weight' (the standard mass) is placed on the left-hand pan, then the balance tips the other way – the left-hand half of the beam goes down and the right hand half goes up – the weight of the standard masses produces an anti-clockwise moment. Balance is obtained in the lever when the clockwise turning effect (or moment) of the weight on one side of the pivot is equal to the anticlockwise turning effect (or moment) on the other side of the pivot.

This is the basic rule governing the balancing of beams, generally called levers. This is an illustration of the 'lever principle' or the law (or principle) of moments which states that

Lever principle

Note that the lever principle applies only if the forces are coplanar.

If a system of coplanar forces keeps a body balanced, the total clockwise moment *about any point in the plane of the forces* is equal to the total anti-clockwise moment about the same point.

Arm of a moment

If we are to use the principle of moments we must know how to calculate the moment of a force about a point. Figure 7.4 shows the line of action of a force F and a point P at a perpendicular distance a from the line of action of the force.

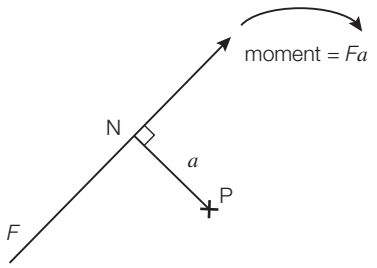


Figure 7.4

The moment of *F* about P is defined as:
(the value of the force) × (the perpendicular distance of P from the line of action of the force)

and so the moment of F about P is $T = F \times a$, where a is called the arm of the moment. Because moment has both magnitude and sense (clockwise or anticlockwise), we must say that the moment of F is $F \times a$ (clockwise), since if one were to pull along the direction of F, the arm PN would turn in a clockwise sense.

We sometimes also speak of the arm of a force that can produce a moment about a point.

The S.I. unit of moment

Since

moment = force × arm

Then

(S.I. unit of moment) = (S.I. unit of force) × (S.I. unit of distance)
= newton × metre
= **newton metre** (N m).

newton metre

Other units of moment that can be used are those based on the sub-units or multiple units of the newton and of the metre. Examples of other units are, for example, N cm, N mm and kN m. The unit used will depend on the size of the moment in question.

Notice that to obtain the S.I. unit of moment we have multiplied a force by the arm which is perpendicular to the line of action of the force. We shall soon be meeting a quite different case where a force is multiplied by a displacement (a length) that is parallel to the line of action of the force, a very different operation, and we shall call that unit the joule, which is the S.I. unit of energy. These two units, the N m (newton metre) and N m (the joule) are not to be confused. They are very different units and are therefore not interchangeable.

Here is a simple example to illustrate the application of the lever principle.

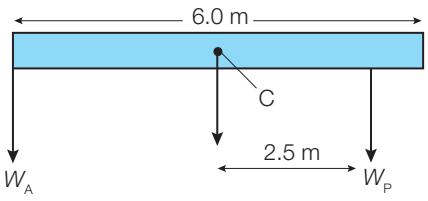


Figure 7.5

Worked example 7.1

A sturdy uniform beam of length 6 m is first balanced at a point directly below its centre of gravity, C, and Allan and his friend, Pierre, use the beam as a see-saw (see figure 7.5). Allan, of mass 50 kg, sits at one end, A, of the see-saw and Pierre finds he can balance Allan when he sits at a point 2.5 m from the pivot on the other side.

(a) (i) Explain why the see-saw balances at a point directly below its centre of gravity.

(ii) List the forces (with their directions) which are acting on the see-saw when it is balanced by Allan and Pierre.

(iii) What is the moment of Allan's weight about the pivot?

(iv) Use the lever principle to calculate Pierre's weight and deduce Pierre's mass.

(b) If the mass of the see-saw is 80 kg, calculate the total force supported by the pivot.

Take g, the gravitational field strength of the earth, to be $10\,\text{N kg}^{-1}$.

Solution

(a) (i) Since the see-saw balances, there can be no net moment acting on it. The only downward force acting on the see-saw when it is first balanced is its weight and the only upward force acting on it then is the reaction at the pivot. Since there is no net moment on the see-saw, and the moment of the reaction is zero, then the moment of the weight is also zero. This is possible only when both lines of action are collinear and this means that the pivot and the C.M. are in the same vertical line. This is an important rule.

(ii) Since the see-saw balances, the forces acting on it must be anti-parallel. We already have two parallel forces in the weights of Allan and Pierre, and at the pivot there must be support for the see-saw in the form of a normal reaction. The forces acting on the see-saw to keep it balanced are therefore (1) Allan's weight downwards, (2) Pierre's weight downwards, and (3) the normal reaction at the pivot (upwards).

(iii) Moment of Allan's weight about the pivot, T_A = force × arm
= Allan's weight × arm
= $mg \times a$ (the formula)
= $50\,\text{kg} \times 10\,\text{N kg}^{-1} \times 3\,\text{m}$
= $1500\,\text{N m}$.

(iv) When the see-saw balances, assuming that Allan is on the left, the moment of Allan's weight (anticlockwise) = moment of Pierre's weight (clockwise)

so $1500\,\text{N m}$ = Pierre's weight × arm of Pierre's weight
= $W_P \times 2.5\,\text{m}$,

where W_P = Pierre's weight

giving $W_P = 600\,\text{N}$.

Since $W = mg$

Then Pierre's mass, $m = W_P/g$
= $600\,\text{N} / 10\,\text{N kg}^{-1}$
= $60\,\text{kg}$

(b) Clearly, the pivot must support the weight of the see-saw, together with the weights of Allan and Pierre.

The total of these weights

= 500 N (for Allan) + 600 N (for Pierre) + 800 N (for the see-saw)

= 1900 N

So the total downward thrust on the pivot = 1900 N.

When a body remains at rest on a pivot or suspended from a support, the C.M. (centre of mass) is in the same vertical line as the pivot or the point of support.

ITQ2

Calculate the moment of a force of 0.05 N acting with an arm of length 50 cm, (a) in N cm and (b) in N m.

Worked example 7.1 involved two downward forces and one upward force. Let us look at an example involving two upward forces and one downward force. Again we notice that there must be anti-parallel forces in the system for equilibrium to be possible.

Worked example 7.2

A uniform lath AB of length 80 cm and mass 50 g is pivoted at a point P, 25 cm from one end (figure 7.6). It is kept horizontal by supporting it at the other end by a vertical string. Calculate the tension in the string and the reaction at the pivot.

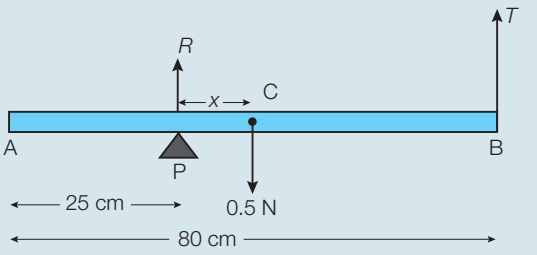


Figure 7.6

Solution

Important points to note here are that:

- the forces are coplanar, and
- the lath is at rest and is horizontal, and so
- the principle of moments *can* be used.

As the lever principle states, any point in the plane of the forces can be taken as the point about which the moments are calculated. Thus, any of the points A, P, C and B can (in theory) be used for the moments equation. It is not a good idea, however, to use just any one of the points A, C or B, because either there will be more than two moments involved if any of these points is used (why use three when you need use only two!) or they will require a knowledge of at least two unknown forces in one equation. We therefore choose a point such that we will have only one unknown force in the moments equation. That point is clearly P, the position of the pivot. With this point as turning point, there will be only two moments to be taken. Two forces will therefore be involved, one of which, the weight, W, is known. There will only be one unknown force, the tension, T, in the string. It is useful to remember that tensions pull. The tension is therefore an upward force.

In taking moments about P, the two arms are as follows:

- arm PC of W, the weight of the rod, where
  PC = 40 – 25 = 15 cm, and
- the arm PB of T, the tension, where
  PB = 80 – 25 = 55 cm

The weight of the lath is given by

$W = mg = \frac{50\,\text{kg}}{1000} \times 10\,\text{N kg}^{-1} = 0.50\,\text{N}$ (remember that in this relation mass, m, is in kg)

Using the law of moments in taking moments about P, we have

moment of W clockwise = moment of T anticlockwise

that is, $W \times (\text{arm of } W) = T \times (\text{arm of } T)$

or $0.50\,\text{N} \times 15\,\text{cm} = T \times 55\,\text{cm}$

giving $T = \frac{0.50\,\text{N} \times 15\,\text{cm}}{55\,\text{cm}}$

$= 0.136\,\text{N}$

$= 0.14\,\text{N}$ (2 sig. fig., since the general significance of the data is two figures)

So the tension in the string is 0.14 N (2 sig. fig.).

ITQ3

If a light newton-meter were used along the string at B in Worked example 7.2, what would it read? Why must the newton-meter be 'light' to give a reliable reading?

ITQ4

Using the data of Worked example 7.2, calculate the value of *R*, the reaction of the pivot by taking moments about C.

Worked example 7.3

Take moments about A (one end of the rod) for the arrangement in figure 7.6 and use the data of the solution to obtain the value of R, the normal reaction at the pivot.

Solution

Since the rod is balanced, taking moments about A,

clockwise moment of W = (anticlockwise moment of R) + (anticlockwise moment of T)

$W \times (\text{arm AC}) = R \times (\text{arm AP}) + T \times (\text{arm AB})$

or $0.50\,\text{N} \times 40\,\text{cm} = R \times 25\,\text{cm} + 0.136\,\text{N} \times 80\,\text{cm}$

so $20.0\,\text{N cm} = R \times 25\,\text{cm} + 10.9\,\text{N cm}$

$R \times 25\,\text{cm} = 20.0\,\text{N cm} - 10.9\,\text{N cm}$

$R = \frac{20.0\,\text{N cm} - 10.9\,\text{N cm}}{25\,\text{cm}}$

$R = \frac{9.1\,\text{N cm}}{25\,\text{cm}}$

$R = 0.364\,\text{N} = 0.36\,\text{N}$ (2 sig. fig.)

In using a value calculated before, we use the 'raw' value 0.136 cm and not the corrected value, 0.14 cm.

Note that we have worked with up to 3 sig. fig and finally corrected to 2 sig. fig.

ITQ5

Now calculate the value of *R* by taking moments about B. Do you get the same result?

You must remember always that:

(i) as long as the lever balances:
sum of clockwise moments = sum of anticlockwise moments

(ii) as long as the C.M. remains at rest:
sum of upward forces = sum of downward forces

ITQ6

A steelyard (figure 7.7) is an old type of weighing machine. It works like a hand-held beam balance. An object is balanced by a known weight, which can be moved along a graduated bar. A hand of bananas was weighed on a steelyard. Use the values given in figure 7.7 to calculate the weight of the bananas.

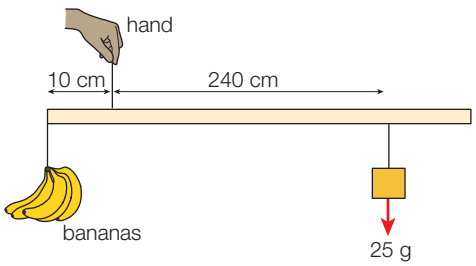


Figure 7.7

Practical activity 7.1

The resultant of two parallel forces

We can show by experiment that the resultant of two or more parallel forces is the sum of the forces. In this activity we will use only two forces. The aim will be:

(i) To test the principle of the lever to see whether it holds in the experiment.

(ii) To show that two parallel downward forces can be replaced by one force which is equal to the sum of the two downward forces.

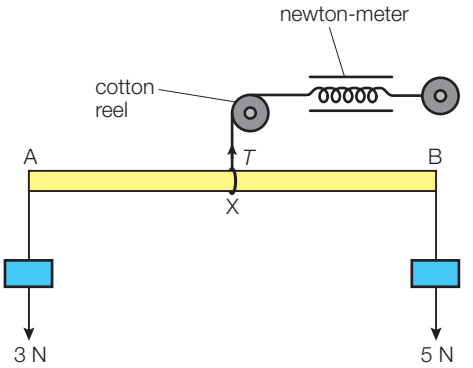


Figure 7.8

You will need:

- two masses of 300 g and 500 g
- a very light but rigid strip of wood of cross-section about 4 mm × 3 mm (the wood used for making kite frames is very suitable)
- a length of thread
- a newton-meter of range 0–10 N × 0.1 N
- two retort stands with clamps
- an empty cotton reel mounted on a sturdy horizontal axis (like a smooth pencil of circular cross-section) held in the clamp of a retort stand
- half-metre rule.

Method

1 Attach the eye at the zero end of the newton-meter to the clamp-rod of one retort stand and the hook to a length of thread with a loop at the free end.

2 Now pass this thread over the cotton reel supported on the pencil as axle, this itself supported on the clamp rod of the other retort stand (see figure 7.8). Adjust the newton-meter so that it is at right angles to the axle of the reel and is also horizontal.

3 Pass the strip of wood through the loop of the thread and suspend the two masses at the ends.

4 Move the strip to and fro in the loop and check for balance on release until this is obtained.

5 Record the reading of the newton-meter at balance and also the distance of the hook from the end A of the strip.

6 Repeat steps 4 and 5 four more times and record the values on a table like table 7.1.

Table 7.1

Observation no.	Distance of hook from end A, *d*/mm	Reading of newton-meter, *T*/N
1		
2		
3		
4		
5		
Mean (or average)		

Using the readings

When the strip was balanced:

1 the clockwise moment (about the end A) of the weight at end B of the rod = the anticlockwise moment of the tension shown by the meter about A;

When the hook of the meter remained at rest:

2 the tension shown by the newton-meter (upward) = the total downward force due to the weights.

So we now have two things to investigate:

(a) Whether (the average tension) × (the mean distance of the hook from A) = (the weight at B) × (the length of the strip).

(b) How close is the reading of the newton-meter to the total of the two weights?

This is a good opportunity to use the limits of error of the reading of the newton-meter to see whether the total weight of the masses might lie within these limits. You should also be able to use the reading of the tension to an appropriate degree of significance to see whether it is very likely that the total weight of the masses might be within the limits suggested by the significance used.

Conclusions

If the equation (a) is found to be true (within the limits of error), then the lever principle is verified, and if (b) is verified, we can then conclude that the two weights can be replaced one equal to the tension in the newton-meter, and this one weight has to be the equivalent of the two and it therefore represents the resultant.

Precautions and discussion

1 The masses must not be allowed to slip along the strip either between or during observations. How will you avoid this? How would slipping affect the observations?

2 Proper balance will have to be ascertained. (How will you do this?)

3 The newton-meter must be horizontal. (Why?)

4 The cotton reel axle must be horizontal. (Why? And how would you ensure this?)

These are good questions for class discussion.

It is also instructive to point out that the position of the hook at balance, X, is the position of the C.G. of the rod and masses. Why is this so?

Finding the position of a centre of mass

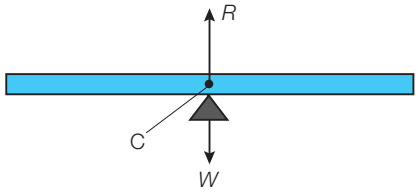


Figure 7.9

Figure 7.9 shows a rod, not necessarily of uniform density or cross-section, balanced on a pivot vertically below its centre of gravity, C. The forces acting on the rod are (i) the weight, *W*, of the rod and (ii) the normal reaction, *R*, of the pivot. The two lines of action of *W* and *R* are the same (the forces are collinear) and so, since this common line of action passes through the pivot, neither force, weight nor normal reaction, has an arm. Neither therefore has a moment. If no moment acts on the rod about the pivot, the rod will remain horizontal and at rest.

It follows then that as long as a body in the shape of a lath or a bar is pivoted and balances on the pivot, the centre of mass of the body will be vertically above the pivot.

For a similar reason, if a body is hung from any point on an object, it will come to rest with the centre of mass vertically below the point of suspension.

Note that this rule holds even if the body is not of uniform density or cross-section.

Having established this principle, we can now apply it to the case cited above. Where must the lever with the masses hanging at the ends be pivoted for balance?

What if the rod is not light (or 'massless')?

Since no object is quite 'massless' (i.e. has no mass at all), this calculation is never really applicable to real-life situations. So what can we do to find the centre of mass of the system, or the point along the rod where the system will balance. We take an example to illustrate what is to be done.

I.T.Q.7

Masses of 500 g and 700 g are attached at the ends of a light rod 24 cm long. By taking moments about one end of the rod, determine where along the rod a pivot should be placed if the rod is to be balanced when supported on the pivot.

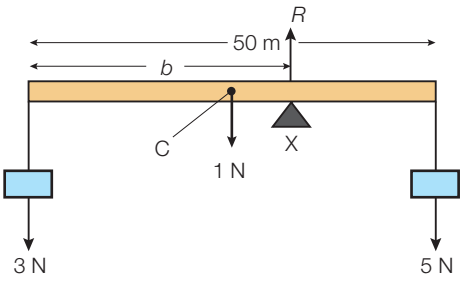


Figure 7.10

Worked example 7.4

A uniform rod of length 50 cm carries two masses of 300 g and 500 g respectively at its ends, A and B. If the mass of the rod is 100 g, where must the rod be pivoted so that it balances? See figure 7.10. This is the same question as: where will the C.G. of the system be situated?

Solution

The C.G. of the system will be that point on the rod where it balances on a pivot. Suppose it balances at the point X, b cm from the end A. Taking moments about the pivot, we have

$$3\,\text{N} \times b + 1\,\text{N}\,(b - 25) = 5\,\text{N}\,(50 - b)$$

or $$3b + b - 25 = 250 - 5b$$

which means that

$$9b = 275$$

giving $$b = 30.55\ldots$$

or $$b = 30.6\,\text{cm}$$

So the system will balance at a point on the rod which is just a little less than 30.6 cm from the end A or about 5.6 cm to the right of the mid-point of the rod. The centre of mass of the system is somewhere along the vertical line passing through this point.

We could make use of the principle that when a body is freely pivoted or freely suspended, its centre of mass will always be above the pivot (if pivoted) or below the point of suspension (if suspended). We make use of this principle in the following activity.

Practical activity 7.2

Locating the centre of mass of an irregularly shaped body of uniform thickness (called a lamina)

Aim

To find the centre of mass of a lamina.

You will need:

- a lamina (thin sheet of stiff material such as plywood or cardboard)
- a length of string with a bob attached
- a small smooth nail (why should the nail be smooth?)
- retort stand and clamp.

Method

1 Make three small, neat holes near the edge of the lamina roughly equally spaced and half the way round. The holes should be large enough to take the nail.

2 Place the head of the nail in the clamp of the retort stand, tighten the clamp, making sure that the nail is horizontal. There should be at least 1 cm of the nail protruding from the clamp.

3 Hang the lamina on the nail, using one of the three holes, and allow it to swing freely until it comes to rest (see figure 7.11).

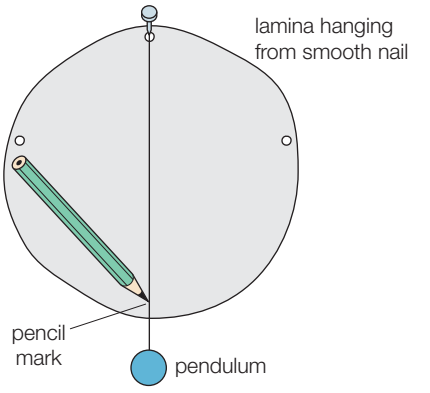


Figure 7.11 Front view of the arrangement.

Precautions

Make sure that friction between the surface of the nail and the material of the lamina is reduced to a minimum (use sandpaper if necessary). Make sure that the hole in the lamina is large enough to accommodate the nail, but not too large!

4 Make a loop at the free end of the pendulum and slip the loop over the projecting end of the nail so that the pendulum almost touches the lamina.

5 Make a pencil mark on the lamina just behind the string, as far as possible away from the hole. (Why as far as possible from the hole?)

6 Remove the lamina from its support. Join the hole just used to the pencil mark just made.

7 Repeat steps 3 to 6 using the second hole round the lamina.

8 Repeat steps 3 to 6 using the third hole.

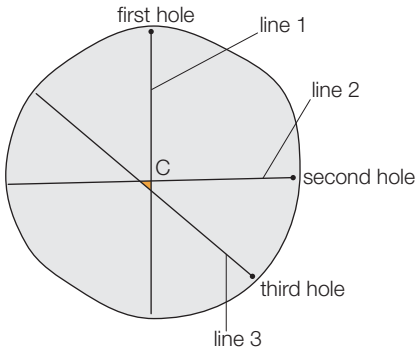


Figure 7.12

Conclusion

The centre of mass of the lamina is the point of intersection of the first two lines drawn on the disc. The third line can be drawn to provide a check for accuracy (figure 7.12). If all three lines meet at one point, the experiment is accurate. If they do not, the accuracy may be judged by the area of the triangle formed by the three intersecting lines at C. The larger the area, the poorer is the accuracy.

This method may be used to find the centre of mass of a lamina of any shape, whether or not the density of the material is the same throughout. What if the thickness of the lamina is not the same all over? Will this method be valid? Discuss.

Having found the position of the centre of mass, how would you check to see if your C.M. is reasonably satisfactory?

The lever principle applied in machines

There are lots of devices that we use daily in the kitchen, in the garage and elsewhere which use the lever principle for their action. We would ordinarily refer to these devices as tools or machines. Whether we say machine or tool, as long as the use of an effort produces a moment about a fulcrum which gives rise to an output force that overcomes a load, we really are talking about levers. Once they use this principle, they are referred to as **levers** and the devices which use the lever principle can be given the general term 'machines'.

lever ›

A machine may be said to be any device, simple or complex, which makes it easier or more convenient to overcome resistive forces (called loads) in doing a specific job. For instance, I would use a pair of shears as a tool or as a machine to cut a sheet of metal. In using the shears I have to use a force on the handles to produce another force on the metal before the metal can be cut. In the context of a machine or a tool, the force I use to get the job done is called the **effort** and the resistance of the metal to being cut is referred to as the **load**. Clearly before the metal can be cut, the force exerted on the metal as a result of my effort must be equal to the load and so we can say that for the job of cutting the metal to be done, the effort must be able to produce a force which is at least equal to the load. Another name for the effort is the **input force** and another name for the force which overcomes the load is the **output force**. It will be clear that the **fulcrum** (another word for pivot) of the lever is the axis about which the blades of the shears turn.

effort › load ›

input force ›
output force ›
fulcrum ›

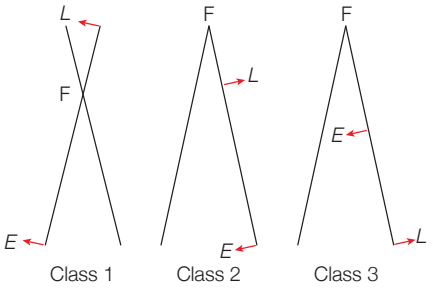


Figure 7.14 The three classes of lever system.

Look at figure 7.13, which shows a pair of scissors, a lever with which we are all very familiar. To cut a piece of cardboard between the blades, you have to press the handles together so that the blade ST turns anticlockwise, while the blade UV turns clockwise. When this happens, blade ST presses on one side of the cardboard, while blade UV presses on the other side. Because forces often occur in pairs and opposite to each other, the cardboard reacts on the blades and exerts opposite forces on them. The force acting on blade ST due to the reaction of the cardboard is *L* in figure 7.13. Force *L* acts to turn the blade ST clockwise about the pivot, P. Force *E* acts to turn it anticlockwise. We therefore have opposite moments acting on the blade about the pivot or fulcrum, P. Note that these are parallel forces. We could therefore apply the law of moments, as we have done before, to the blade ST.

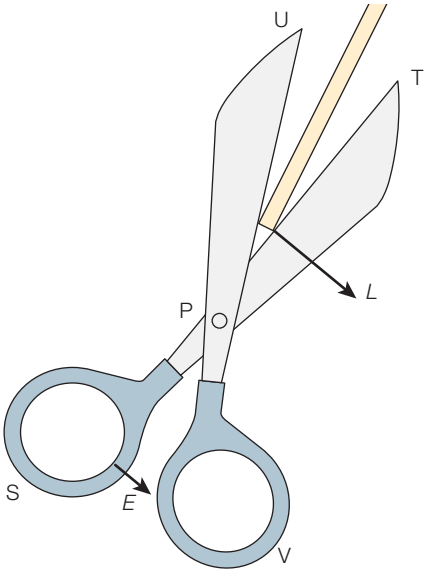


Figure 7.13 The forces acting on one blade of a pair of scissors. Similar forces will act on the other blade.

lever law ›

Where one applied force gives rise to another, we refer to the law of moments as the '**lever law**' or the 'lever principle'.

The fulcrum of a lever will not always be placed between the effort and the load (as in the case of the scissors). In some cases the load is between the fulcrum and the effort (as in nut-crackers), and in others the effort is between the fulcrum and the load (as in tongs or tweezers). Levers are therefore divided into three classes, depending on where the fulcrum is in relation to the effort and the load. These three classes and their features are summarised in table 7.2 and figure 7.14. Some examples are shown in figure 7.15.

Table 7.2 The three classes of levers

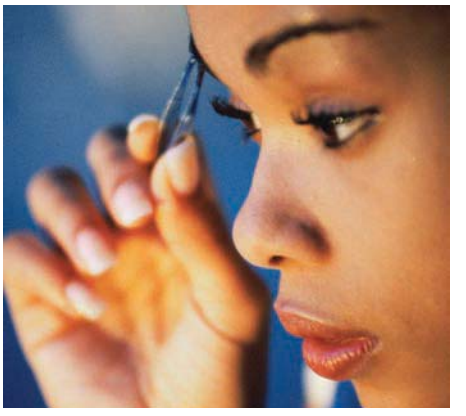
Class	1	2	3
Features	F between *E* and *L* F usually nearer to *L* than to *E*	*L* between F and *E* F nearer to *L* than to *E*	*E* between F and *L* F nearer to *E* than to *L*
Examples	see-saw, claw-hammer, scissors, crow-bar	wheelbarrow, nut-cracker, can opener	broom, pick-axe, tweezers, tongs, fishing rod, garden fork



(a)



(b)



(c)

Figure 7.15 Levers at work: (a) class 1, (b) class 2, (c) class 3.

Mechanical advantage

In table 7.2, F represents the fulcrum of the lever, E the force or effort applied to 'do the job' and L the load, which is the output force produced that overcomes the resistance when the effort is applied.

The benefit of some levers lies in the fact that quite often the force to be used is rather less that the value of the load being overcome. For example, we can use a claw-hammer (like that shown in figure 7.15 (a)) whose handle is always very much longer than the distance between the butt end (or heel) of the hammer and the 'V' of the claw. This gives an advantage to the user of the hammer and so we say that the hammer offers to the user a **mechanical advantage** (M.A. for short), mechanical because the forces involved are all mechanical and the benefit constitutes an advantage to the user. To understand the significance and meaning of the advantage, consider the line diagram of figure 7.16 in which a claw-hammer is being used to draw a nail out of a piece of wood. The user applies an effort E at right angles to the handle of length a. This effort gives rise to an output force (equal to the load L) which will be used to overcome the resistance of the wood. Assuming that the frictional resistance of the wood is R and that it is equal to the load, L, we can use the lever law and write that

effort × arm of effort = load × arm of load

or $$E \times a = L \times b$$

See figure 7.16.

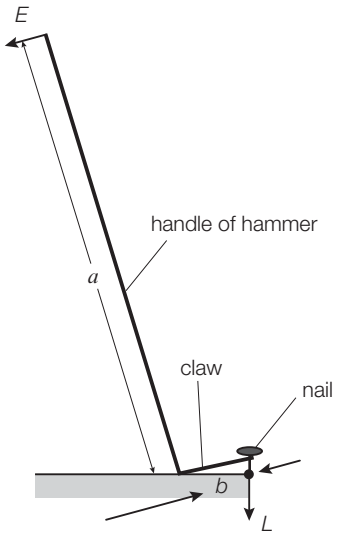


Figure 7.16

This equation gives the mechanical advantage, by definition, load/effort, as

$$\text{M.A.} = \frac{L}{E} = \frac{a}{b}$$

(which is much larger than 1, since the handle is always so long – the longer the handle the better!).

You will be able to conclude from this that the theory of the crow-bar as a machine is very similar to that of the hammer. In the case of the crow-bar, with the much longer handle, it is possible to obtain a much better M.A. and so the effort needed to overcome a given load is made smaller.

An alternative way of looking at the matter above, is to observe that the ratio $E/L = b/a$ and so

$$E = (b/a) \times L$$

which shows that the effort used to draw the nail is much smaller than the frictional resistance of the wood to the nail, (b/a) being a small ratio.

A further mathematical point is to observe that since $E = (b/a) \times L$

Then $E = (b \times L)/a$

For a hammer with a fixed value of b and a given job to be done (L constant) the effort needed varies inversely as the length of the handle. So if the hammer handle is too short to do the job, then the solution could be to lengthen it in some way (by using a length of steel pipe?) since, by doubling the handle length, we should need only half the effort. Of course, this is only true if the machine is ideal or 100% efficient. The question of efficiency is discussed in chapter 9.

You will note that from the information given in the table 7.2 above, the M.A. of a lever used as a machine could be either greater than 1 or less than 1. See figure 7.15 for examples of levers whose mechanical advantages are shown in table 7.3.

ITQ8

The handle of a claw-hammer is 30 cm and the distance from the butt (heel) to the 'V' of the claw is 3 cm. What is the effort needed to draw the nail from a piece of hardwood whose frictional resistance to the nail is 150 N?

MATHEMATICS: inverse variation

Table 7.3 Classes of lever.

Class of lever	1	2	3
Values of M.A.	greater than 1	greater than 1	less than 1
Examples	crow-bar, claw-hammer	nut-cracker, wheelbarrow	long-handled broom, garden fork

Stability

Sometimes if we accidentally knock against a vase of flowers it does not return to rest but topples over. Why? We will find that moments and the position of the centre of mass play an important part.

Consider an object resting on a horizontal surface. Figure 7.17 illustrates the effect of tilting an object whose base is getting increasingly smaller and smaller, but whose centre of mass is always the same height above the base. As the base reduces in size, the line of action of the weight of the object moves closer to the edge about which the body is tilted.

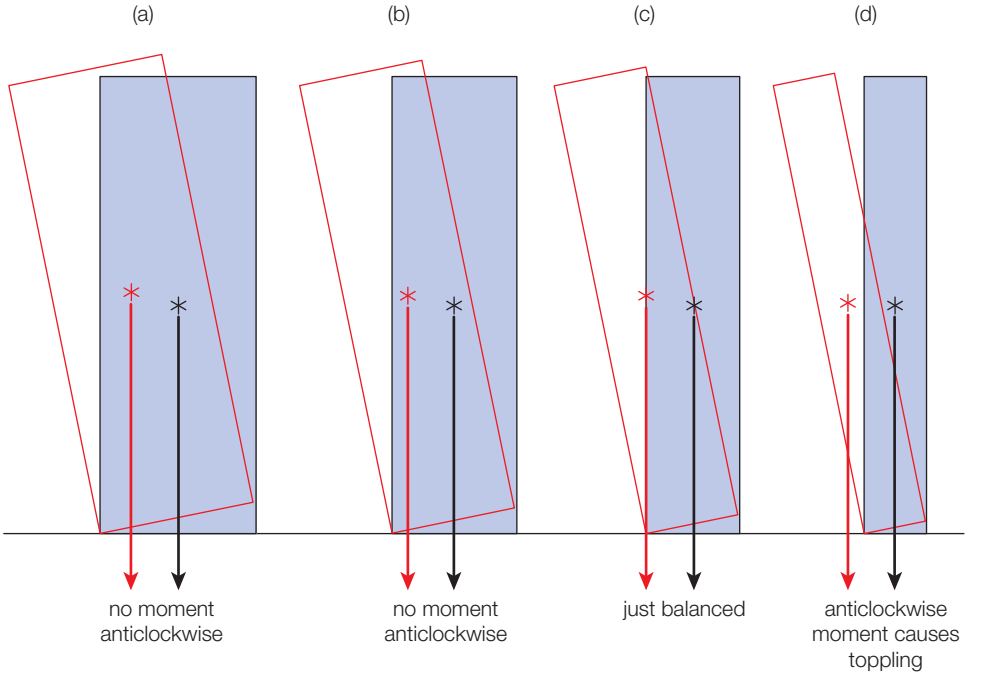


Figure 7.17 Toppling of tall objects.

As long as the line of action of the weight remains within the limits of the base, there is a moment exerted by the weight that tends to turn the body clockwise. This clockwise moment will have the effect of putting the body back to rest. There will be no toppling.

stable ›

At positions (a) and (b) the body is able to return to rest, and so it is said to be **stable**. At (c), where the line of action just passes through the turning point, the body is on the point of toppling. At (d), where the line of action of the weight passes outside the turning point, the moment becomes anticlockwise and the body topples. When it is at position (d), it is **unstable**, because, although it is at rest while being held, it will not return to rest when it is released. Therefore, 'stable' means 'able to return to rest after being slightly displaced from rest through a small angle'.

unstable ›

Factors that increase stability are the width of the base and the height of the centre of mass above the base. Figure 7.18 (a) shows that an object with the

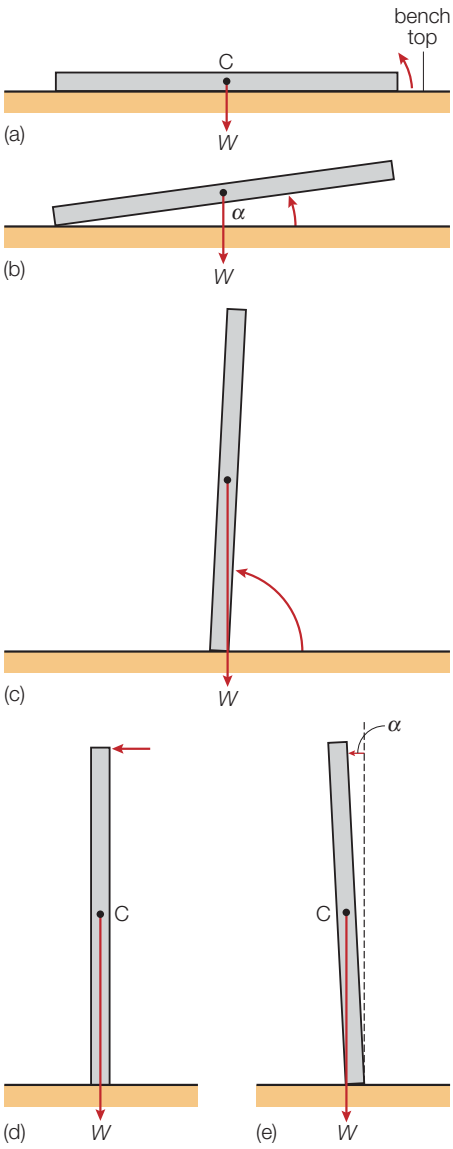


Figure 7.19

same base size is stable when the centre of mass is low but not when the centre of mass is above a certain height. If the centre of mass is low and the object is tilted, the moment of the weight is clockwise and the object returns to its original position. If the centre of mass is high, then even a small tilt results in the turning moment of the weight being anticlockwise and the object topples. Figure 7.18 (b) shows why an object with a wide base is stable, even though the C.M. is high and why one with a narrow base is not.

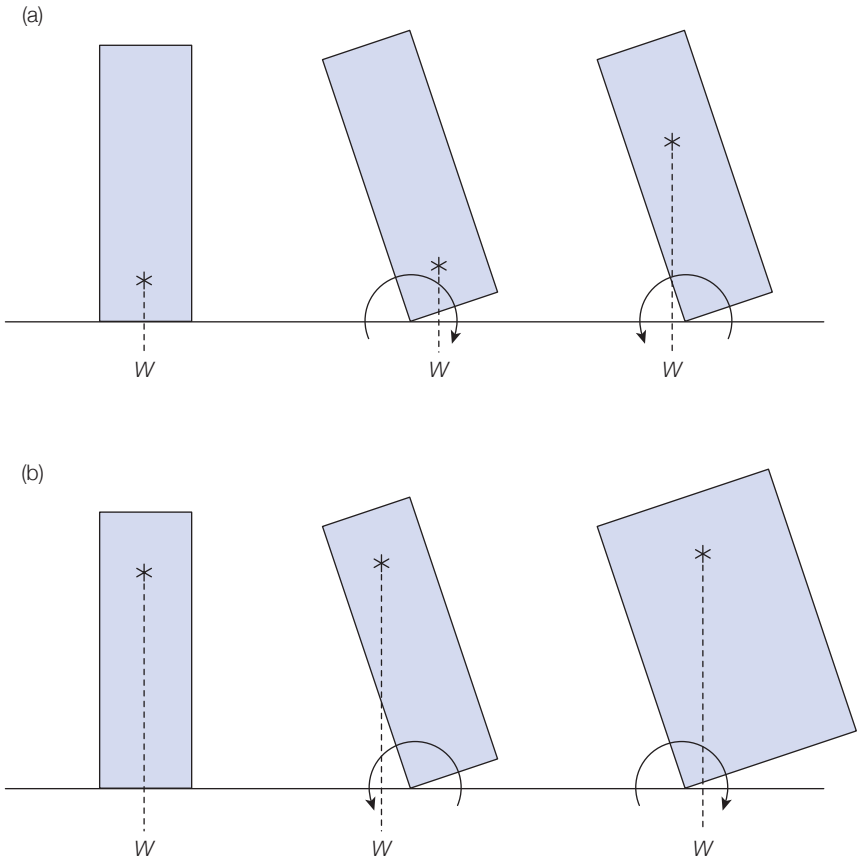


Figure 7.18 The factors that affect toppling are the position of the centre of mass and the size of the base. A curved arrow shows the direction in which turning takes place. In (a) as the C.M. gets higher the body topples earlier; in (b) as the base gets wider the body topples later.

ITQ9

Use the diagrams of figures 7.19 (a)–(e) to explain why the rod in the position labelled (a) is much more stable than the same rod in position labelled (d)

Chapter summary

- Coplanar forces are forces whose lines of action are all in the same plane.
- Coplanar forces may be parallel, anti-parallel or non-parallel.
- Anti-parallel forces are forces with parallel lines of action, but two opposite directions.
- Coplanar parallel forces always have a resultant.
- Coplanar anti-parallel forces may or may not have a resultant.
  - If they have a resultant, the body on which the forces act will move in the direction of the resultant.
  - If they do not have a resultant, the centre of mass of the body on which they act will remain at rest.
- Rotation is caused by the turning effect of a force, called a moment.
- The moment is the product of the force and the perpendicular distance from the turning point to the line of action of the force. This perpendicular distance is called the arm of the moment.

- If a body remains at rest under anti-parallel forces, then the resultant of the forces acting in one direction balances the resultant of those acting in the opposite direction.
- In equilibrium, the total clockwise moments of forces about any point in the plane of the forces will balance the total anticlockwise moments.
- The stability of a body is a measure of the ability of the body to maintain its equilibrium when it is displaced to one side through a small angle.
- The stability of a body depends on:
  - the height of its centre of mass above the base;
  - the width of its base.
- For good stability:
  - the centre of mass should be as low as possible, and
  - the base should be as wide as possible.

Answers to ITQs

ITQ2 **(a)** 2.5 N cm **(b)** 0.025 Nm

ITQ3 0.14 N. If the newton-meter is not light it will read the tension as well as its own weight, thus making the reading of the tension inaccurate.

ITQ4 $R = \frac{(0.136\,\text{N} \times 40)}{15} = 0.36\,\text{N}$

ITQ5 $R = \frac{(0.5\,\text{N} \times 40)}{55} = 0.36\,\text{N}$

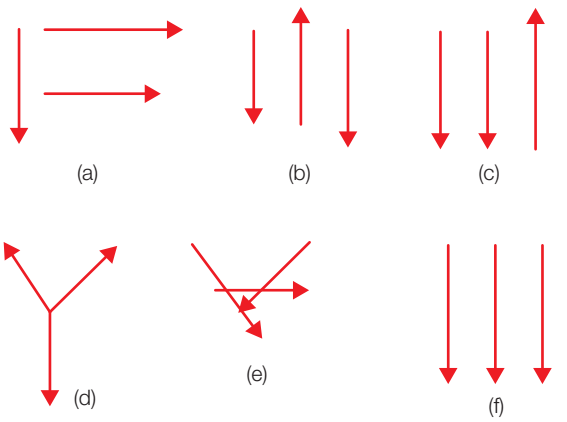
ITQ6 6 N

ITQ7 14 cm from the 500 g mass

ITQ8 15 N

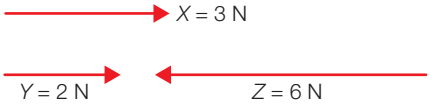
Examination-style questions

1 (i) State whether each of the set of forces shown is parallel, anti-parallel or non-parallel.



(ii) State whether equilibrium is possible or not possible in each of the set of forces shown. Give a reason for each answer.

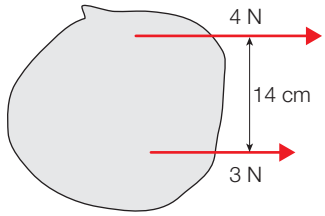
2 (i) State the magnitude of the resultant of the forces shown.



(ii) Show the direction of this resultant.

(iii) Show the magnitude and direction of the force that will produce equilibrium when put together with this resultant.

3 Parallel forces of 4 N and 3 N act on a sheet of cardboard separated by a distance of 14 cm.



(i) What is the magnitude of the resultant of these forces?

(ii) Where between the two given forces must a balancing force be applied to make the resultant force on the sheet of cardboard zero?

(iii) If the resultant force did become zero what would happen to the sheet?

4 (i) State Hooke's law.

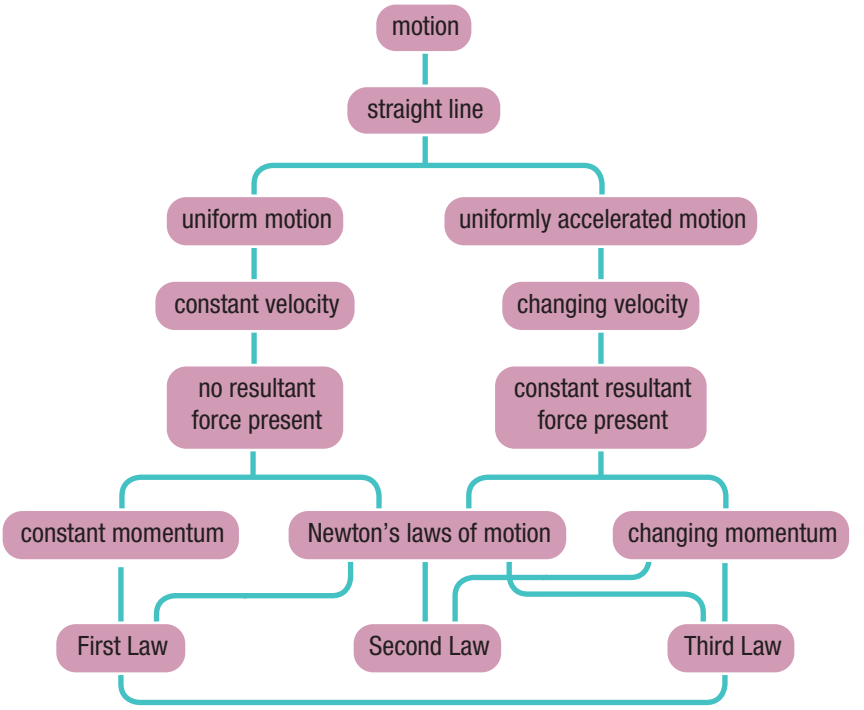
(ii) Give an example to illustrate the meaning of Hooke's law.

(iii) A certain spring requires an initial force of 5 N to separate the coils before the spring begins to stretch. The spring then stretches 1 cm for every 12 N of load applied. If the initial length of the spring is 40 cm, how long will the spring be when the load applied to it is 48 N?

8 Motion in a Straight Line

By the end of this chapter you should be able to:

- define the terms distance, displacement, speed, velocity and acceleration
- state Aristotle's argument in support of his hypothesis that the speed of an object varies directly as the force acting on the object
- state Newton's three laws of motion
- use Newton's laws of motion to solve problems on motion in a straight line
- understand the difference between distance–time and displacement–time graphs
- understand the difference between speed–time and velocity–time graphs
- use displacement–time and velocity–time graphs to solve problems
- define linear momentum as the product of mass and linear velocity
- describe situations which demonstrate the laws of motion
- apply the laws of motion to solve problems
- use Newton's laws of motion to explain the working of common dynamical systems



Galileo and Newton

Motion is one of the most common experiences of everyday life, but the subject did not engage the attention of many scientists until the 16th century. Aristotle, the Greek philosopher mentioned in chapter 4, proposed that the velocity of an object depended on its speed – that a constant force applied to an object produced a constant speed in the object. He based this hypothesis on the observation that the harder the horses pulled a chariot, the faster the speed

of the chariot became, but no attempt was made to support this argument by experiment. Nor were any attempts made to measure either the pulling force of the horses or the speed of the chariots. His hypothesis was based merely on conjecture.

It was not until around the end of that century, with the invention of reliable methods of measuring time, that serious study of motion was undertaken. This subject interested two of the most eminent scientists in the 16th and 17th centuries: Galileo Galilei (whose work was summarised in chapter 4) and Isaac Newton, of whom we shall learn more later in this chapter. Galileo did not concern himself too much with the 'why' of motion. He was more interested in the 'how'. In 1638 he wrote:

CHAPTER 4

> *it is the purpose of our Author* [referring to himself] *merely to investigate and to demonstrate some of the properties of accelerated motion (whatever the cause of this acceleration may be).*

Remembering the work of Galileo

It was Galileo, an Italian mathematics professor who, by studying the motion of spheres rolled down slopes and of bodies falling vertically through large distances, first studied the characteristics of straight-line motion experimentally. Through his experiments he was able to work out how straight-line motion took place.

We will therefore (like Galileo) be looking into the question of motion in a straight line which takes place under constant acceleration. Note that in this treatment:

- the motion will always be in a straight line, and
- the acceleration will be constant.

First, however, we must define the terms that we will be using in describing straight line motion:

distance travelled ›
- **Distance travelled**. This is the length of the path followed by a moving object. It is a scalar quantity and is therefore taken without any regard to direction of movement.

displacement ›
- **Displacement**. This is the distance, measured in a straight line, from the starting point to the position of a moving object at any moment. It is always measured from the origin or starting point of the motion, and is a very different quantity from distance travelled, as we shall soon see.

speed ›
- **Speed**. This is the rate at which distance is being covered irrespective of direction.

velocity ›
- **Velocity**. This is the rate at which displacement is changing with time. Like displacement, it is a vector quantity and therefore has direction.

acceleration ›
- **Acceleration**. This is the rate at which a velocity or a speed is changing with time. Acceleration can be seen either as a vector or as a scalar, depending on whether the change that gave rise to it was a change of velocity or a change of speed.

The relationship between 'distance–time' graphs and speed

Imagine two cars, A and B travelling on a straight road. A is travelling at a constant speed of $20\,m\,s^{-1}$ to the right and B is also travelling at a constant speed of $20\,m\,s^{-1}$, but to the left. We will begin by imagining that at time $t = 0\,s$ car A is just passing a point O on the road, the origin of the motion, travelling at $20\,m\,s^{-1}$ to the right. The other car, B, is not yet in the picture, but is on the road travelling to the left at the same speed of $20\,m\,s^{-1}$.

It takes car A 5 seconds to reach point X on the road 100 m away from O, the origin (see figure 8.1 (a)). Table 8.1 shows the distance, s, between the origin and car A and the time, t, after the start. The graph of figure 8.1 (b) shows the graphical relationship between the distance travelled by car A and time.

Table 8.1

Time after the start, *t*/s	Distance travelled in this time, *s*/m, from O
0	0
1	20
2	40
3	60
4	80
5	100

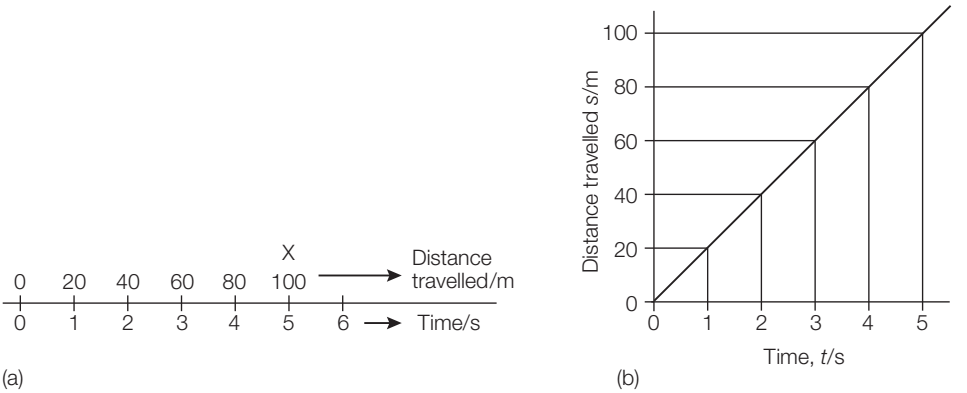


Figure 8.1

The graph of s against t is a straight line. This graph is shown in figure 8.1 (b). Notice that the slope, defined as $\frac{\text{(change of distance travelled)}}{\text{(time taken for the change)}}$

$$= \frac{\text{(final distance – initial distance)}}{\text{(time for distance to change)}}$$

$$= \frac{(100\,\text{m} - 0\,\text{m})}{(5\,\text{s} - 0\,\text{s})}$$

$$= 20\,\text{m s}^{-1}$$

speed ›

The fact that the slope is constant reflects the fact that the **speed** is constant, which means that the distance covered each second is always the same. The graph does not tell us in what direction the car is travelling. All it needs to show is how much distance the car has covered in unit time, and it does that.

The slope of a distance–time graph represents 'speed'.

'Displacement–time' graph and velocity

Let us turn now to the 'displacement–time' graph. When we plot the distance of the car from the origin (that is, the displacement from O), we get an identical graph to the first one. The slope of this graph will now have the same numerical value, but since we have divided a change of displacement by a time interval, this gives us the **velocity**, which is defined as the rate of change of displacement with time, instead of speed, as before.

velocity ›

So far no distinction has been made between the velocity and the speed of car A. If there was a difference it would be that the speed is $20\,\text{m s}^{-1}$ and nothing more, whereas the velocity is $20\,\text{m s}^{-1}$ in an easterly direction (the direction of the motion). Remember that velocity is a vector quantity and its direction must be given to completely describe it. The slope of the displacement–time curve is positive. This tells us that the velocity is in a positive direction, which we already know is to the right.

The slope of a displacement–time graph represents velocity in magnitude and direction.

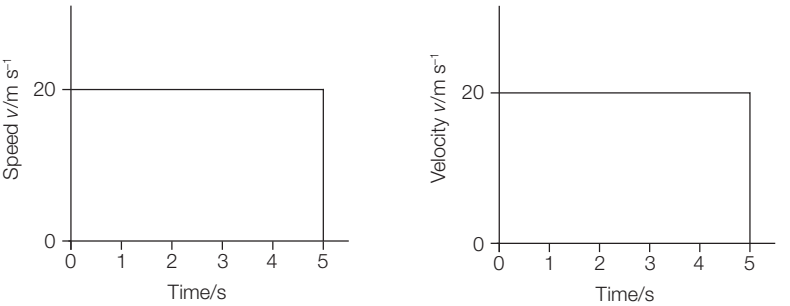


Figure 8.2

speed–time graph ›
velocity–time graph ›

Velocity is a vector.

Speed is a scalar quantity.

The **speed–time graph** for the motion of the car is the same as the **velocity–time graph**. Although the graphs do not reveal it, the velocity of the car is specifically from left to right. The speed–time graph might represent a motion taking place in any number of different directions. The two graphs of figure 8.2 are completely interchangeable. The only thing different is the labelling of the vertical axis – one is labelled 'speed', the other is labelled 'velocity'.

You will have noticed that the slope of a speed–time graph has units $m\,s^{-1}/s$ or $m\,s^{-2}$. This implies that the acceleration of a motion may be obtained by taking the slope of either a speed–time graph (in which case the acceleration is a scalar) or a velocity–time graph (in which case the acceleration is a vector).

The slope of a speed–time graph or a velocity–time graph represents an acceleration.

Area under a speed–time graph

If we calculate the area enclosed by the speed–time graph, the two ordinates, $t = 0\,s$ and $t = 5\,s$, and the time axis (referred to as 'the area under the graph', for short), and using area = length × breadth, we obtain 5 s (for length) × $20\,m\,s^{-1}$ (for breadth) = 100 m.

If we do a similar calculation for any type of motion, whether at constant speed or not, we can find the distance covered. The area under a speed–time graph is the distance covered.

Area under (or above) a velocity–time graph

In similar manner, if we take the area above (or below) a velocity–time graph, we can obtain the displacement that the moving object undergoes.

The area above (or below) a velocity–time graph is the displacement that the moving object undergoes.

Let us now take an example involving an object which, like the car just considered, is moving in a straight line but, unlike the car, is moving with a changing velocity, the velocity increasing by the same amount every second.

uniform acceleration ›

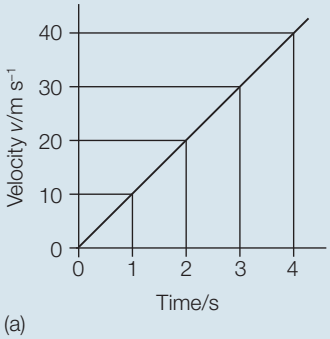
A motion in which the speed or the velocity of an object changes by a constant amount every second is said to be moving with **uniform acceleration** and the motion is referred to as a uniformly accelerated motion. If the velocity is increasing with time, the object is said to be accelerating; if the velocity is decreasing with time, the object is said to be decelerating.

Worked example 8.1

A ball falls from rest under the force of gravity only for 4 s. What would be its velocity at the end of the fall? Take the acceleration due to gravity (which is the rate at which the velocity of a body falling under the action of gravity increases each second) to be $10\,m\,s^{-2}$.

Solution

Saying that the acceleration is $10\,m\,s^{-2}$ means that every second the velocity of the ball increases by $10\,m\,s^{-1}$. So, if the initial velocity is $0\,m\,s^{-1}$, the diagram of figure 8.3 shows that in 4 s the velocity acquired would be $10\,m\,s^{-2} \times 4\,s = 40\,m\,s^{-1}$.



(a)

Figure 8.3

Worked example 8.2

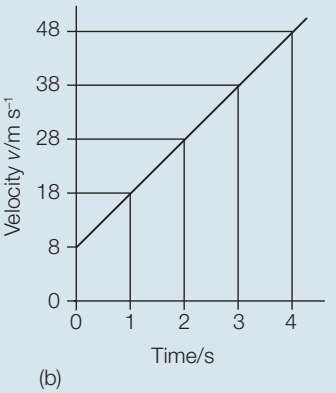
Suppose the ball of Worked example 8.1 was released from a helicopter that was descending vertically at $8\,m\,s^{-1}$.

(i) Draw a velocity–time graph for the motion of the ball.

(ii) What would be the ball's velocity 4 seconds after being released?

Solution

(i)



(b)

Figure 8.4

(ii) As the ball has been released from a moving object, its initial velocity will be that of the moving object. Every second thereafter the velocity will increase by $10\,m\,s^{-1}$ and so after 4 seconds the velocity would be the sum of the initial velocity and the additional velocity due to the acceleration.

After 4 seconds the velocity is given by

v = initial velocity + increase due to acceleration

We write $v = u + at$ (the first equation of motion), where v = the velocity after time, t, commonly called 'the final velocity', u = the initial velocity, a = the constant acceleration and t = the time the motion lasted.

Substituting in the equation for this case, we have the velocity of the ball after 4 seconds,

$$v = 8\,m\,s^{-1} + 10\,m\,s^{-2} \times 4\,s$$
$$= (8 + 40)\,m\,s^{-1}$$
$$= 48\,m\,s^{-1}$$

MATHEMATICS: algebra – $y = mx + c$ or $y = c + mx$

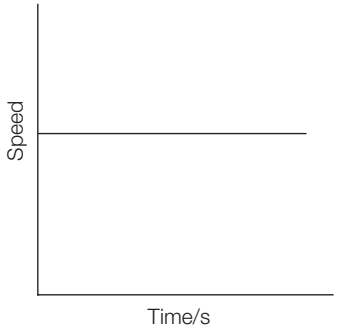


Figure 8.5

ITQ 1

The value of the acceleration due to gravity at the Earth's surface is $10\,m\,s^{-2}$.

(i) How long would it take an initial downward speed to increase from $10\,m\,s^{-1}$ to $40\,m\,s^{-1}$?

(ii) What is the acceleration for the motion represented by figure 8.5?

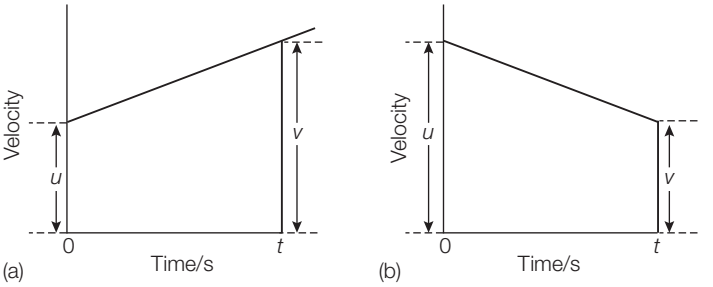


Figure 8.6

The general graph of velocity against time for a uniformly accelerated motion is shown in Figure 8.6 (a). It is a straight line and will always be straight with a

positive slope as long as the acceleration is positive or the velocity is increasing regularly with time (constant acceleration). If the motion was with constantly decreasing velocity, the acceleration would be negative, and the graph would again be a straight line but with a downward slope (see figure 8.6 (b)).

From the first equation of motion, we find that $a = (v - u)/t$, which is what we would expect if acceleration is defined as the change of velocity in unit time. It is important that you stick to the formula

$$\text{acceleration} = \frac{\text{change of velocity}}{\text{time taken for change}} = \frac{\text{final velocity } (v) - \text{initial velocity } (u)}{\text{final moment } (t) - \text{initial moment } (0)}$$

If this formula produces a negative result, it means that the motion is a decelerated one, and that the object is slowing down with time.

MATHEMATICS: algebra – slope of a straight line graph $= (y^2 - y^1)/t$

IT Q2

Sketch a speed–time graph to show each of the following motions:

(i) A ball moving very fast and with constant speed that is suddenly stopped.

(ii) A humming-bird that flies horizontally at constant speed, stops for a while and then flies vertically upwards at twice the speed.

(iii) An ant that moves quickly forwards, stops for a while, then crawls very slowly backwards.

(iv) A ball dropped from a certain height, rebounding twice on the ground at half the speed of impact each time.

The phrase 'acceleration due to gravity' is sometimes referred to as 'the acceleration of free fall'.

Worked example 8.3

A stone, A, is released from a point X, high above the ground, Two seconds later another stone, B, is released from the same point (figure 8.7 (a)).

(i) Draw a velocity–time graph for the motion of each stone.

(ii) Use the graphs to determine how far apart the stones would be 6 seconds after the first stone is released. Take the acceleration of free fall, $g = 10\,\text{m s}^{-2}$.

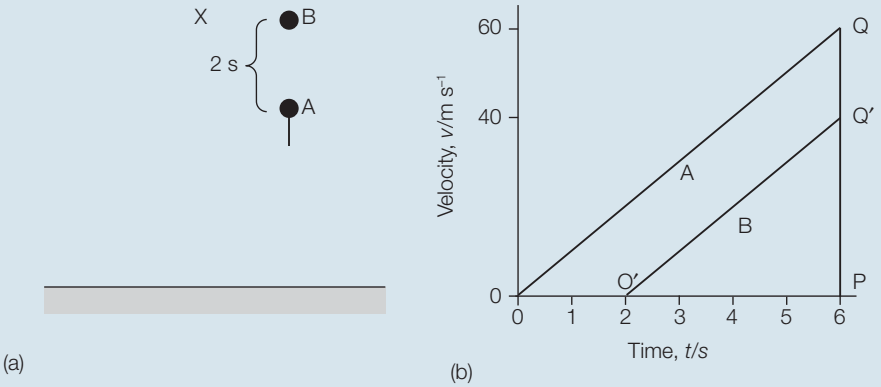


Figure 8.7

Solution

Figure 8.7 (a) shows the position of each stone at the moment of its release.

(i) Figure 8.7 (b) shows the velocity–time graphs for the two motions. Note that the slopes of the two graphs are the same. This must be so, since both stones are subject to the pull of gravity and will therefore have the same increase in velocity per second, or the same acceleration, $10\,\text{m s}^{-2}$.

(ii) The areas under the respective graphs give the displacements of the two stones – see figure 8.7 (b).

The velocity of A after 6 seconds is

$$\begin{aligned} v_A &= u + at \\ &= 0 + 10\,\text{m s}^{-2} \times 6\,\text{s} \\ &= 60\,\text{m s}^{-1} \end{aligned}$$

So the displacement of stone A = the area under the velocity–time graph for stone A or the displacement of stone A from the starting point,

Graphs of motions with the same acceleration will have the same slope.

s_A = the area under graph A

= ($\frac{1}{2}$ base × height) of ΔOPQ

= $\frac{1}{2}$ × 6 s × 60 m s^{-1}

= 180 m

The velocity of stone B after falling for 4 s (2 seconds later than A)

= 4 s × 10 m s^{-2} = 40 m s^{-1}

So the displacement of B from the starting point, X, after 4 s, is

($\frac{1}{2}$ × base × height) of ΔO′PQ′ = $\frac{1}{2}$ × 4 s × 40 m s^{-2}

= 80 m

Since the displacements both have the same direction (downwards from the point of release), the separation of the stones will therefore be given by

$s_A - s_B$ = 180 m – 80 m

= 100 m

The stones will therefore be separated by 100 m.

MATHEMATICS: geometry – area of a triangle = $\frac{1}{2}$ (base) × (height)

ITQ3

A car starts from rest and accelerates at a constant rate. After travelling a distance of 200 m, the speed of the car is 80 km h^{-1}. What was the acceleration of the car?

We will now go back to the case we considered at the start of the chapter concerning cars A and B.

Suppose that at the very moment car A is at the point X travelling to the right, car B passes the same point X in the opposite direction travelling at the same speed of 20 m s^{-1}. As time continues, car B gets closer and closer to the origin, O. Table 8.2 shows how the distance of car B from O (the displacement of B) varies with time measured from the start of the observation.

Table 8.2

Time interval from the start for car B, *t*/s	Distance of car B from the origin, O (displacement of car B), *s*/m
5	100
6	80
7	60
8	40
9	20
10	0

The distance–time graphs (figure 8.8 (a)) for the two cars travelling between O and X will have exactly the same shape, but not the displacement–time graphs (figure 8.8 (b)). The distance graphs must look exactly the same, since the two cars both have the same speed and so will have covered the same distance in the same time. You will have noticed that they both have the same slope – this is no surprise, since we are only concerned with the magnitude or the rates of change of distance, and not the direction.

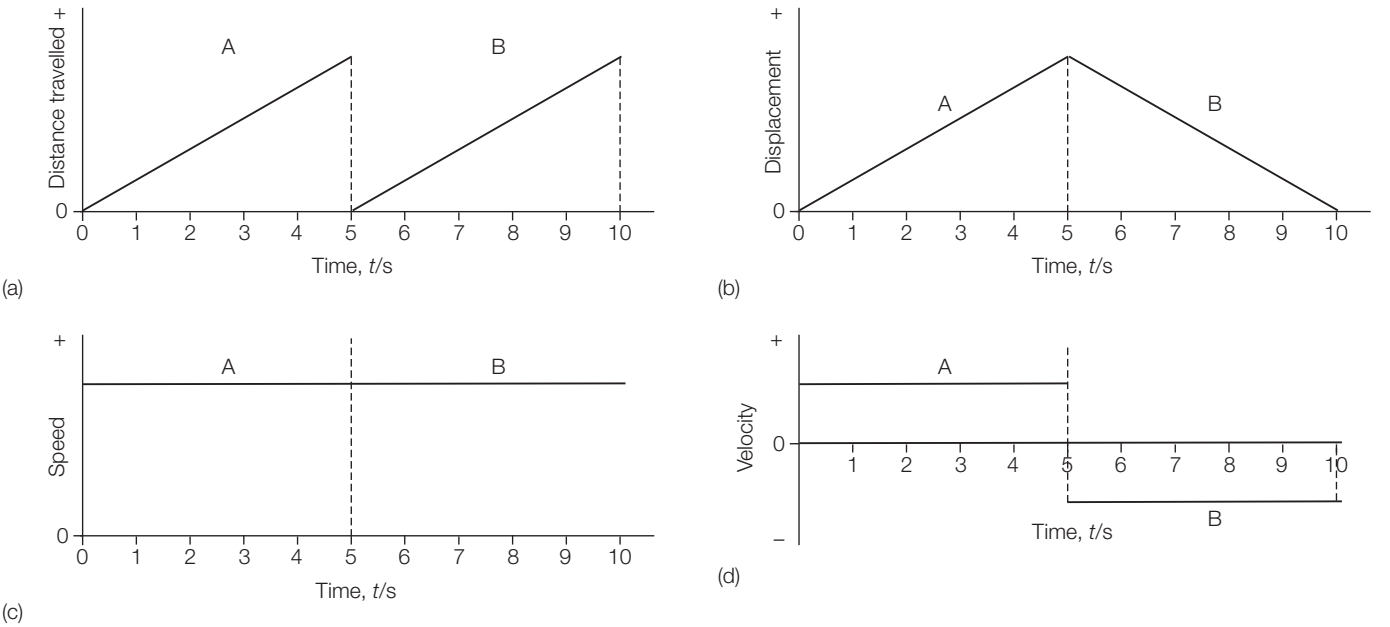


Figure 8.8 Graphs for two cars, A and B: (a) distance–time; (b) displacement–time;(c) speed–time; (d) velocity–time.

Note the difference in the slopes of the two displacement graphs, however. Graph A rises uniformly, showing an increasing displacement as time passes. Graph B falls uniformly, since the displacement (the separation of the car from O) gets less and less with time at a uniform rate. Notice, also, that since the slope of graph A rises, showing a positive rate of change with time, or velocity, the associated velocity is positive. Graph B shows a negative slope, indicating a negative velocity. We expect the velocity of car B to be negative since it is moving to the left starting point, X. This is a good example of the dependency of velocity on displacement and also on direction.

Thus, the slope of graph A, defined as

$$\frac{\text{final displacement} - \text{initial displacement}}{\text{time taken for change}}$$

gives $\frac{(100-0)\text{m}}{(5-0)\text{s}} = +20\,\text{m s}^{-1}$. However, the slope of graph B $= \frac{(0-100)\text{ m}}{(10-5)\text{ s}}$ $= -20\,\text{m s}^{-1}$.

Note that the displacements and the times at which they occur must correspond. Thus the coordinates for A are (0 s, 0 m) initially and (5 s,100 m) finally, and for B (5 s, 100 m) initially and (10 s, 0 m) finally.

You will notice that, although the magnitude of the velocity is the same, $20\,\text{m s}^{-1}$, as we would expect, the sign is negative, indicating that, since velocity is a vector, it is in the opposite direction to the velocity of car A. So the velocity of car B is negative because its motion is to the left. This does not depend on starting point. Displacement does and so the displacement of both cars is always positive (and so above the time axis), as both cars are at all times to the right of the origin, O.

Let us now take a look at the speed–time and the velocity–time graphs for the cars. For car A, the velocity is to the right and so, as is the custom, we will take this velocity as positive. For car A the velocity is positive (since the slope is positive), whereas car B's velocity is going to be negative, as shown by the negative slope of the velocity–time graph and by the motion of B to the left.

Figure 8.8 (c) shows the speed–time graphs for the two cars. These are the same for both cars. When we consider the velocity–time graphs (see figure 8.8(d)) for A, however, the velocity is at a constant value of $+20\,\text{m s}^{-1}$ for the first 5 s, and for B it is a constant value of $-20\,\text{m s}^{-1}$ for the second 5 s. Note that when we multiply the constant positive velocity of $+20\,\text{m s}^{-1}$ by the time of travel, 5 s, we are really multiplying a constant velocity by a time and getting a displacement of +100 m to the right from A's starting point. When we do the same for B using its starting point, X, we obtain a negative displacement to the left of X. When we add together these two displacements, one relative to O and the other relative to X, we obtain +100 m + (–100 m) = 0, showing that the overall displacement of cars A and B is zero. Remember that both displacements were taken relative to the origin, O. Compare this addition of two vectors with that of the addition done in figure 5.1 (b).

This study of motion is called **kinematics**. What we have done applies to motion in a straight line only. It has not taken into account the factors or conditions that produce a change in velocity. Why does a body move as it does? Under what conditions will it speed up or slow down? Why does a body move in a circle? A study of these matters is a study of Newton's laws. It is a study of **dynamics** to which we now turn.

ITQ4

An object is projected vertically up into the air from a balcony high above the ground. It comes to rest momentarily after 3 seconds and then returns to a point 10 m below its point of projection at a time t_0 seconds later.
For the entire motion, draw: (a) a velocity–time graph; (b) a speed–time graph; and (c) an acceleration-time graph which shows that acceleration can be considered a vector quantity. Show the quantities given where possible on your graphs.

MATHEMATICS: adding vectors

kinematics ›

dynamics ›

Dynamics and Newton's laws of motion

Newton's contribution to physics and mathematics

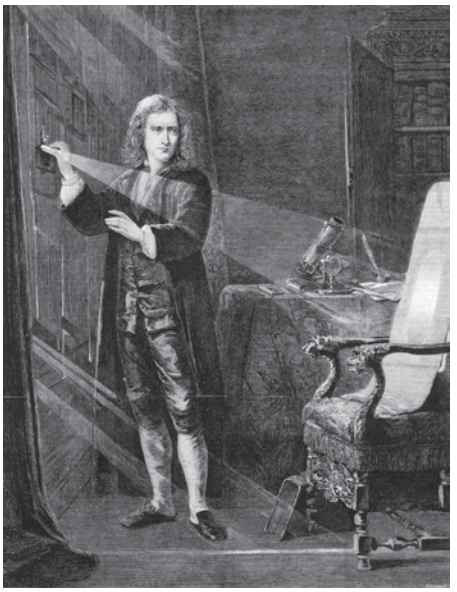


Figure 8.9 Sir Isaac Newton.

The laws of motion are named after Newton (figure 8.9) because he was the first to show an understanding of the relationship between force and mass. In 1686 he published his *Principia Mathematica*, upon which the principles of the subject now known as 'dynamics' are based. It is unlikely that he carried out experiments to prove his theories of momentum, force and gravity, but the hypotheses he put forward were based on mathematics (most of which he developed himself to develop his theories) and so could be verified by observation of natural phenomena such as the movement of the planets.

Newton studied this movement with the use of a telescope he designed himself – called the 'Newtonian' telescope – which formed the image in the telescope by means of a set of mirrors instead of lenses.

The fact that Newton's name is associated with experiments in other branches of physics shows that his interests in physical phenomena were many and varied. These interests ranged from optics (Newton's rings due to interference), to heat (Newton's law of cooling by convection), to the collision of spheres (Newton's coefficient of restitution equation), and to the dispersion and diffraction of light.

Internet search terms: Newton's rings/ Newton's law of cooling/Newton's coefficient of restitution equation

We will now investigate dynamics in a straight line, which is essentially a study of Newton's three laws of motion.

These laws, formulated by Isaac Newton in his *Principia*, form the basis of what is now called 'mechanics'. The laws give the rules governing the behaviour of matter (mass) when it is subjected to forces. It was Newton who first proposed that the behaviour of matter is controlled by forces of two kinds, *mechanical* and *gravitational*.

Did you know that Newton is also credited with being one of the two mathematicians who first developed what is now known as the calculus?

Newton proposed three laws. Each of the laws is concerned with motion and with force.

Newton's First Law

Newton's First Law of motion ›

Newton's First Law states that, for the motion of an object to change, some external force must act on that object. We say:

> **An object at rest will remain at rest or continue in a state of uniform motion in a straight line unless acted on by an external force.**

ITQ5

Use Newton's First Law of motion to explain why a bicycle needs to be pedalled even when travelling at constant speed along a level road.

We can look at this law in a different way. Everyday experience tells us that it is harder to stop a heavy stone rolling down a hill than to stop a light one. It is harder to stop a cyclist moving at 40 km h^{-1} than to stop the same cyclist moving at only 10 km h^{-1}. So both mass and velocity are involved, which leads us to the idea of **momentum**.

momentum, *p* ›

> **The momentum of a body is the product of its mass and its velocity:**
> **momentum = mass (a scalar) × velocity (a vector)**
> **Therefore momentum is also a vector: $p = m \times v$**

The plural of momentum is momenta.

Momentum is a vector quantity. The direction of the momentum of a body is the same as the direction of the velocity of the body. Its units are the units of mass and velocity.

S.I. unit of momentum

Don't forget that momentum is a vector quantity. Its symbol is *p*.

kilogram metre per second ›

S.I. unit of momentum = S.I. unit of mass × S.I. unit of velocity
= (kilogram) × (metres per second)
= $\mathrm{kg\,m\,s^{-1}}$

Another way of stating Newton's First Law is to say that:

> **As long as there is no external force acting on a system, the momentum of that system remains constant (is conserved).**

conservation of momentum ›

Momentum is a vector quantity. So when we say that momentum is conserved, we mean that neither the value (mass × velocity) nor the direction of the momentum changes. The **conservation of momentum** is embodied in what is commonly known as Newton's first law of motion. So if momentum is conserved in a collision of two vehicles, it means that:

> **total momentum before collision = total momentum after collision**

Newton's Second Law

Newton's Second Law of motion ›

Newton's Second Law connects the force acting on a body to its acceleration. From the statement of the Second Law, it can be shown that, if a force acts on an object, then

> **force = (mass of the body) × (acceleration of the body)**
> **or $F = m \times a$**

This is related to the rate of change of the momentum. If an object accelerates from velocity u to velocity v, its change in momentum is given by

(final momentum) – (initial momentum), or

$mv - mu = m(v - u)$

If this has happened in time t, then

$$\text{rate of change of momentum} = \frac{\text{change in momentum}}{\text{time}} = m\,\frac{(v-u)}{t}$$

But $\frac{(v-u)}{t} = a$, the acceleration, so

$$\text{rate of change of momentum} = \frac{\text{change in momentum}}{\text{time}} = m \times a$$

provided that the units used for force, momentum and time all belong to the same system of units. This system is, of course, S.I.

In other words, whenever the equation, force = mass × acceleration, is used, the force must be in newtons, the mass in kilograms, the velocity in metres per second and the time in seconds.

So, if force = $\frac{\text{change in momentum}}{\text{time}}$

then, on cross multiplying, we have

force × time = change in momentum

Inserting the units for each of these quantities, we have

newtons × seconds = $\mathrm{kg\,m\,s^{-1}}$

which shows that instead of using '1 $\mathrm{kg\,m\,s^{-1}}$', we could use '1 N s' instead. Therefore:

> **The unit of momentum, 1 $\mathrm{kg\,m\,s^{-1}}$ = 1 N s (read '1 newton second').**

IT Q6

A mass of 10 kg, initially at rest, is acted on by a force of 10 N for 10 s. (i) What is the final velocity of the mass? (ii) In which direction is it moving?

IT Q7

A recap! Define the newton.

Newton's Third Law



Newton's Third Law of motion ›

This law shows us, for example, why a rocket moves forward as gas shoots out behind it, and why a car stops moving when it hits a tree (figure 8.10). The law says that if a body A exerts a force on another body B, then body B exerts an equal and opposite force on body A. You will remember that we met the concept of 'paired forces' in chapter 5. This concept is in keeping with **Newton's Third Law**, which may be briefly expressed as follows:

> **To every action there is an equal and opposite reaction.**

A quick summary of Newton's laws is shown in table 8.3.

Table 8.3 A summary of Newton's laws of motion.

Law	Deals with	Resultant force
1st	uniform motion in a straight line *or* equilibrium	no external force is present
2nd	accelerated motion	external force acts
3rd	uniform and accelerated motion	external forces may be present but internal forces cancel





Figure 8.10 Action and reaction.

You might argue that motion cannot occur if action and reaction are equal and opposite, because their resultant is zero. But consider the following case.

If I pull on a rope that is tied around a tree, there might be a state of equilibrium between my hand that pulls the rope, the rope and the tree. I pull on the rope with a certain force, P, and the rope reacts to my pull by pulling back with a tension, T, which (according to the Third Law) is equal to my pull, P, but opposite to it. These forces, P and T, act on different bodies: the tension of the rope acts on my hand and the muscles in my hand act on the rope.

Note that these two equal forces act on different bodies. This is what the Third Law stipulates. Put another way, the law states that

> **To every action there is an equal and opposite reaction, these two, the action and the reaction, acting on different bodies.**

To demonstrate that the rope pulls my hand, we may cut the rope. What happens then? I fall backwards. Why? Because when my hand was at rest holding on to the rope, we had a state of balanced forces: the pull of the rope on my hand and the pull of the muscles in my hand on the rope. This could be seen as a validation of Newton's First Law (no unbalanced force, and therefore no momentum change – the momentum of my hand continues to be zero). There is also the case for Newton's Third Law to be argued, since my pull on the rope (the action) has given rise to a tension in the rope (the reaction). These two act on different bodies, one, the tension of the rope acting on my hand, and the other, the pull of my muscles, acting on the rope.

When the rope is cut, my hand is no longer under the action of two equal and opposite forces, but under the action of only one force, that due to the muscles in my arm. This is where the Second Law comes into play. There is now no longer a tension to 'balance' the force in my arm. The result is that my hand moves in the direction of the now 'unbalanced' force, the force in my arm, which is backwards.

There are many other examples that might be cited where the principles of at least one of the laws is demonstrated. Some of the more common situations are:



With the help of Newton's Second and Third Laws describe how a trampoline can slow down an athlete who lands on it in a vertical direction.



Which do you think is better to avoid being injured: to use a stiff trampoline to break your fall or to use a soft one? State clearly the reasons for your choice.

- The behaviour of billiard balls when struck by a cue (Laws 1 and 3).
- The firing of a firearm and its recoil (Laws 1 and 3).
- The action of the jet engine of an aircraft or a spacecraft (Laws 1 and 3).
- The obvious recoil of a fire hose (Laws 1 and 3).
- The action of a garden sprinkler (Laws 1 and 3).
- Alighting from a moving vehicle (such as a train) (Law 1).
- Releasing air from a blown-up balloon (Laws 1 and 3).
- Putting a screw into a wall with a screwdriver while standing on a smooth floor (which law?).
- A short pendulum in a car inclines itself away from the vertical while the car accelerates or decelerates (which law(s)?).

What happens to the pendulum when the car moves at constant speed in a straight line?

It might well be argued that the application of the Second Law is present, but very subtly, in all of the cases mentioned above, since when the force of the Third Law comes into existence it will cause the object on which it acts to experience an acceleration. This is often manifested in a 'jerk' experienced by the object.

Worked example 8.4

A frictionless trolley, A, of mass 400 g travelling along a straight, level track at 5 m s^{-1} collides with another frictionless, stationary trolley, B, of mass 600 g (figure 8.11). On impact, they stick together and move along with the same speed. Calculate the speed of the two trolleys just after they collide.



Figure 8.11

Solution

Since the trolleys are frictionless, there is no external force due to friction acting on them. The track is horizontal and so the effect of gravity is zero. We use Newton's First Law, which says that, with no external force acting on the trolleys:

momentum before collision = momentum after collision

So we have

momentum before collision = (mass of A) × (initial velocity of A)
= 400 g × 5 m s^{-1}
= 2000 g m s^{-1}

because trolley B has zero velocity and it therefore has zero momentum. This value of 2000 g m s^{-1} is therefore equal to the momentum after collision. Let the final joint velocity be v. Then

momentum after collision = (mass of A + B) × (velocity of A + B)
= (400 + 600) g × v

So

(400 + 600) g × v = 2000 g m s^{-1}

giving

$$v = \frac{2000\,\text{g m s}^{-1}}{1000\,\text{g}} = 2\,\text{m s}^{-1}$$

Worked example 8.5

A ball, A, of mass 2 kg moving along a straight line at $4\,m\,s^{-1}$ collides with another ball, B, of mass 4 kg moving along the same line, but in the opposite direction, at $6\,m\,s^{-1}$. On colliding with ball B, ball A moves backwards at $5\,m\,s^{-1}$. Calculate the velocity of ball B after the collision in magnitude and direction.

Solution

Total initial momentum of the two balls before collision

$= 2\,kg \times 4\,m\,s^{-1}$ (for A) $+ 4\,kg \times -6\,m\,s^{-1}$ (for B)

Here we are taking vectors acting to the right as positive and those acting to the left as negative.

So the total momentum before collision

$= 8\,kg\,m\,s^{-1} + (-24\,kg\,m\,s^{-1}) = -16\,kg\,m\,s^{-1}$

Since there was no external force acting on the two balls together during the impact, the total momentum after they collide must not change from the original value.

The total momentum after the collision $= 2\,kg \times -5\,m\,s^{-1} + 4\,kg \times v$, where v is assumed to be the unknown velocity of B after the collision.

So

$-16\,kg\,m\,s^{-1} = -10\,kg\,m\,s^{-1} + 4\,kg \times v$

or

$-16\,kg\,m\,s^{-1} + 10\,kg\,m\,s^{-1} = 4\,kg \times v$

Giving

$4\,kg \times v = -6\,kg\,m\,s^{-1}$

and

$v = -1.5\,m\,s^{-1}$

The negative sign of the velocity shows that the velocity of ball B after the collision is $1.5\,m\,s^{-1}$ in the same direction (to the left) which it had before the collision.

Worked examples 8.4 and 8.5 illustrate the application of Newton's First Law. We now look at examples of the application of the Second and Third Laws.

Worked example 8.6

The collision of the balls in the last example lasted 0.1 second. Calculate (i) the change in momentum of each ball; (ii) the rate of change of the momentum of each ball; and (iii) the force acting on each ball to change its momentum.

Solutions

(i) Change in momentum = final momentum – initial momentum

For ball A:

change $= 2\,kg\ (-5 - 4)\,m\,s^{-1} = 2 \times -9\,kg\,m\,s^{-1} = -18\,kg\,m\,s^{-1}$

For ball B:

change $= 4\,kg\ (-1.5 - (-6))\,m\,s^{-1} = 4 \times 4.5\,kg\,m\,s^{-1} = +18\,kg\,m\,s^{-1}$

Have you noticed that the change in momentum of ball A = –(change in ball B)? Does this not mean that the overall change in momentum for both balls = 0? This is Newton's First Law in action!

(ii) Rate of change of momentum = (change in momentum)/time

For ball A:
rate of change = $-18\,\text{kg m s}^{-1}/0.1\,\text{s} = -180\,\text{kg m s}^{-2}$

For ball B:
rate of change = $+18\,\text{kg m s}^{-1}/0.1\,\text{s} = +180\,\text{kg m s}^{-2}$

newton ›

Since $1\,\text{kg m s}^{-2}$ has been given the name **newton**, then the rates of change of momentum for the balls (which is equal to the force acting on each ball) is the same for both in magnitude.

(iii) Since the signs of these changes are opposite, then the forces acting on the two balls are opposite in direction. This is an illustration of the Third Law.

What we just did was not a proof of the laws, but a verification of them. We will now discuss investigations that can be carried out to test the laws in order to find out to what extent they are true.

Testing Newton's laws by experiment

Newton's First and Second Laws can be tested using a device called a ticker-timer and a tape. The ticker-timer is a device that makes it possible to measure small times with a precision of about 0.02 second. The tape is used to measure the distance moved by the object used in the experiment. Essentially the device measures distance travelled by a moving tape and the ticker is used to determine time intervals.

How the ticker-timer device works

The ticker-timer measures time by making a dot at intervals of 0.02 s (for a 50 Hz current supply) on a length of tape attached to the moving object. Figure 8.12 shows the arrangement.

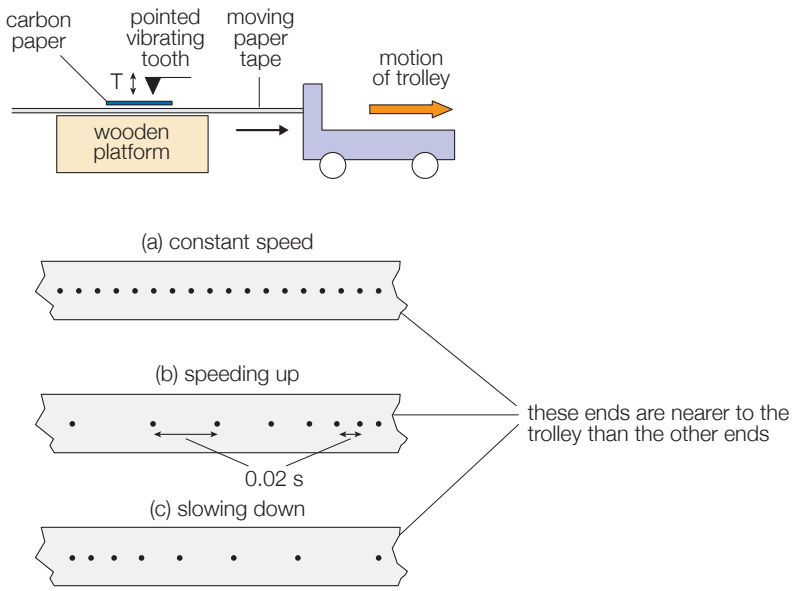


Figure 8.12 The use of a ticker-timer.

It your supply is 60 Hz, the time interval is 0.017 s – a really awkward value to deal with!

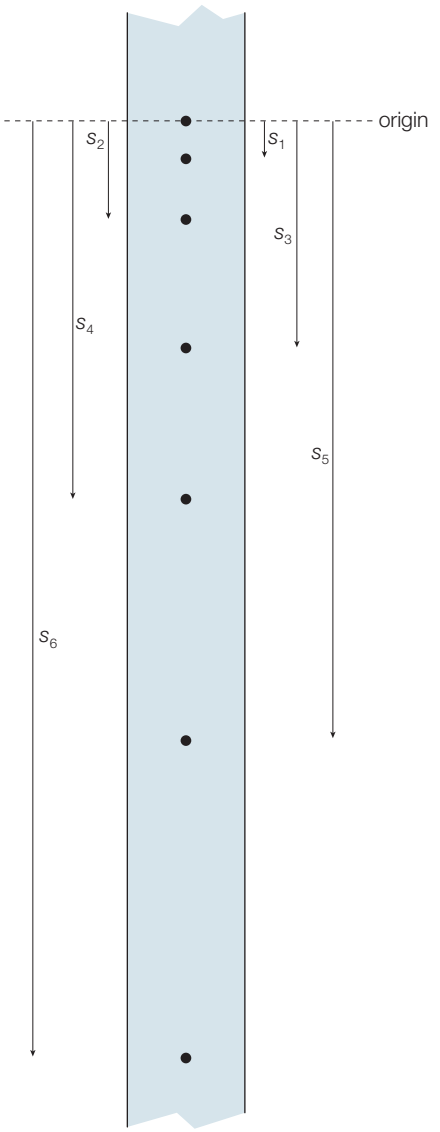


Figure 8.13 A length of timer tape.

A length of white tape, attached to a trolley with 'frictionless wheels', is drawn by a trolley between a vibrating tooth and a wooden platform. This tooth is operated by an electromagnet attached to a 50 Hz or 60 Hz a.c. electricity supply. For every cycle of current, the pointed tooth descends once on to the tape under it. Since there is a disc of carbon paper lying face downwards above the tape, when the tooth descends, it takes the disc of carbon paper with it and a black dot is left on the tape at the point where the tooth touches it. If the tooth descends 50 times (say) each second, it descends every 1/50th of a second. This means that the dots will be separated by a time interval of 0.02 s.

Clearly, how far apart two successive dots are will depend on the speed at which the paper is moving past the tooth. At the end of an observation, therefore, there would be a row of dots on the tape, equidistant if the tape was moving at a constant speed, getting closer if it was slowing down, but further and further apart if it was speeding up. The appearance of the tape in these three cases is shown in figure 8.13.

Measurement of acceleration with a ticker-timer and tape

It can be proved by considering the area under a velocity–time graph that, in a uniformly accelerated motion,

$$s = ut + \tfrac{1}{2}at^2$$

where the symbols have their usual meanings. If we divide throughout by the time, t, we get an equation for the average speed, v^*, where

$$v^* = \tfrac{s}{t} = u + \tfrac{1}{2}at$$

If a graph is plotted of $\frac{s}{t}$ against t, and it is straight, this indicates that the motion was uniformly accelerated, since a must be a constant for the graph to be a straight line.

A length of timer tape and the measurements derived from it are shown in figure 8.13 and table 8.4.

We can tell how the acceleration is changing in the case of a non-uniformly accelerated motion by following the shape of the average speed–time curve. In figure 8.14 the lines show motions in which the average speeds are changing uniformly (indicating uniform rate of change of velocity). In figure 8.15, however, the curves show non-uniformly accelerated motions. Graph A shows fairly uniform speed followed by a steadily increasing acceleration. Graph B shows at first a motion with constant deceleration followed by further steadily increasing deceleration to zero speed.

Table 8.4 Measurements from the ticker tape.

Distance, s/cm	Time, t/ticks	Average speed /cm per tick
0.5	1	0.50
1.3	2	0.65
3.0	3	1.00
5.0	4	1.25
8.2	5	1.64
12.4	6	2.10

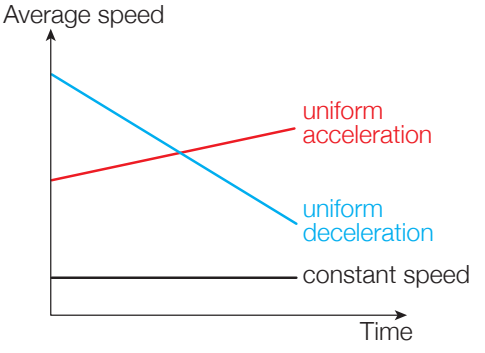


Figure 8.14

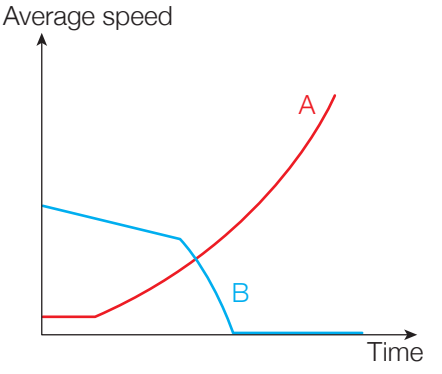


Figure 8.15

Practical activity 8.1

Testing Newton's First Law

You will need:

- runway
- block
- two trolleys
- ticker-timer and tape
- pin
- cork
- set of masses.

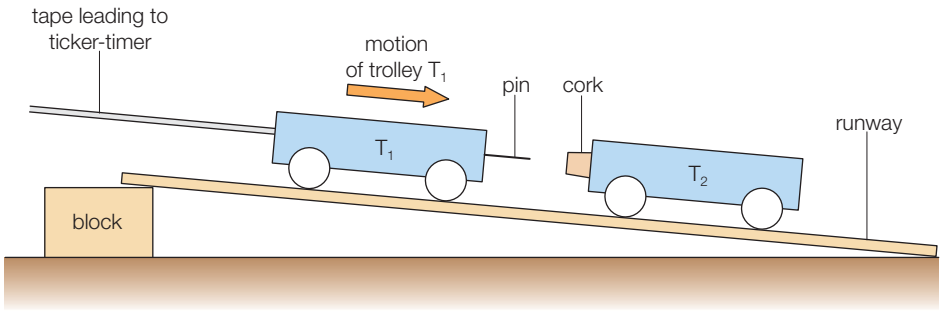


Figure 8.16 Experimental arrangement for testing Newton's First Law.

Method

1 Set up the apparatus as shown in figure 8.16. Incline the runway to compensate for friction.

2 Place one trolley (T_1) about one-quarter of the way down the runway away from the end nearer the timer. Attach a length of tape to the trolley and thread it through the timer.

3 Place the other trolley (T_2) about half-way down the runway so that it is in line with trolley T_1. Trolley T_2 should remain at rest if the runway is properly inclined for friction compensation.

4 Now give trolley T_1 a gentle push down the incline. It will run down the incline and collide with trolley T_2. Together they will move down the runway, stuck to each other.

5 Tear off the length of tape with the dots. The dots should show two sets of different spacings. The first set (set A) should be more widely spaced than the second set (set B). Measure the total length of tape occupied by 10 spaces in each of the two sets of dots. Denote the length of 10 spaces in set A by d_A and the length of 10 spaces in set B by d_B.

6 Record the masses of the trolleys m_1 and m_2, the lengths d_A and d_B and the ratios d_A/d_B and $(m_1 + m_2)/m_1$ as shown in table 8.5.

7 Add a small mass (of about 200 g) to trolley T_2 to increase the value of m_2. For the new mass m_2, repeat steps 2 to 6. Record the values of the quantities indicated at step 6 above in table 8.5.

8 Add further masses of 400 g, 600 g and 800 g to m_2, and for each new value of m_2 repeat steps 2 to 7.

Table 8.5 Table of results.

m_1/g	m_2/g	$(m_1 + m_2)$/g	d_A/cm	d_B/cm	$\frac{(m_1 + m_2)}{m_1}$	$\frac{d_A}{d_B}$

Theory

If Newton's First Law holds, then momentum is conserved:

$$m_1 \times v_A = (m_1 + m_2) \times v_B$$

But

$$\text{velocity} = \frac{\text{displacement}}{\text{time}}$$

so we can write the equation for conservation of momentum as

$$\frac{(m_1 \times d_A)}{t} = (m_1 + m_2) \times \frac{d_B}{t}$$

Multiplying each side by t gives

$$m_1 \times d_A = (m_1 + m_2) \times d_B$$

or

$$\frac{(m_1 + m_2)}{m_1} = \frac{d_A}{d_B}$$

The theory predicts that the ratio of the total masses of trolleys T_1 and T_2 to the mass of trolley T_1 should be equal to the inverse ratio of the distances covered by the masses in the same time.

Results

When table 8.5 is completed, the values obtained in the last two columns of any row should be equal for given values of m_1 and m_2. If this is the case, then the law has held well in the experiment.

As an alternative method of testing the law, a graph of d_A/d_B against $(m_1 + m_2)/m_1$ may be plotted. If the law has held, then the graph should be a straight line of slope 1.

Practical activity 8.2

Testing Newton's Second Law

You will need:

- runway
- block
- trolley (of mass about 500 g)
- ticker-timer and tape
- string
- pulley
- scale pan
- set of masses.

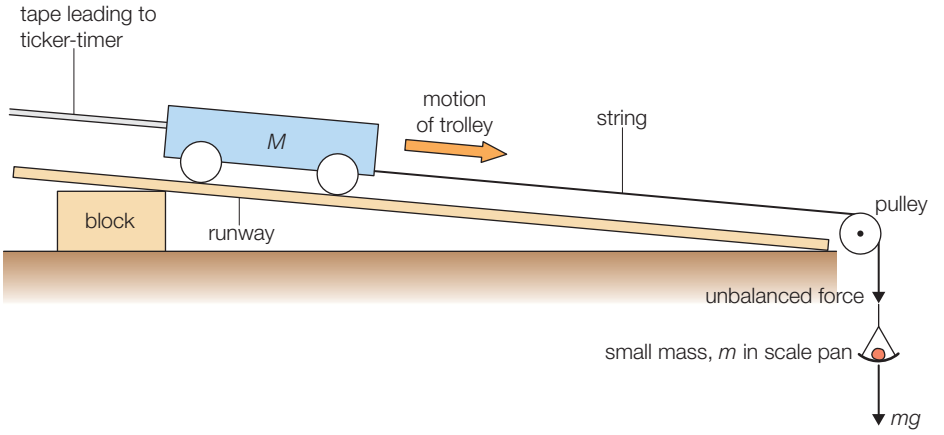


Figure 8.17 Experimental arrangement for testing Newton's Second Law.

Method

1 Set up the arrangement shown in figure 8.17, with the runway inclined to compensate for friction. Record the mass M of the trolley. Put a small mass m (about 10–20 g) in the scale pan. Hold the trolley at rest on the runway.

2 With the tape attached to the trolley, release the trolley and allow it to run a distance of about 1 metre down the runway.

3 Tear off the dotted length of tape, make measurements of distances and the corresponding number of spaces between ticks. Record these values as shown in table 8.6.

4 Plot a graph of these values and from it calculate the acceleration, as in the theory section on page 121.

5 Add a 100 g mass to the trolley (so its total mass M is now about 600 g) and

repeat steps 2 to 4 with the same small mass m in the scale pan.

6 Carry out the same steps 2 to 5 for total trolley mass $M = 700\,g$, $800\,g$ and $900\,g$.

7 Record the values of the total mass of the trolley, M, and the corresponding acceleration, a, in a table (see table 8.7).

Table 8.6 Table for one set of results.

Number of spaces, N	Distance occupied by N spaces, d/cm	Mean speed (d/N), v/cm per tick
5		
10		
15		
20		
25		

Table 8.7

Mass of trolley, M/kg	Acceleration produced, a/m s^{-1}	$M \times a$
0.500		
0.600		
0.700		
0.800		
0.900		

Theory

If Newton's Second Law holds, then for a constant mass m:

unbalanced force $\propto$ acceleration

acceleration $\propto \frac{1}{\text{mass}}$

The force pulling the trolley is provided by the weight, mg, of the mass, m. If m is small compared with the mass M of the trolley, and there is proper compensation for friction, then the unbalanced force is always mg. If

$$s = ut + \tfrac{1}{2}at^2$$

where t = the number of ticks ($\propto$ time interval) for the journey, then the average velocity, v^*, is

$$v^* = \frac{s}{t} = u + \tfrac{1}{2}at$$

However, $u = 0$. So a graph of average velocity, s/t, against t should yield a straight line (figure 8.18). The slope of this line $= (\frac{1}{2})a$, where a = acceleration. The acceleration, a, is 2 × slope. The units will be cm per tick^2 (cm tick^{-2}).

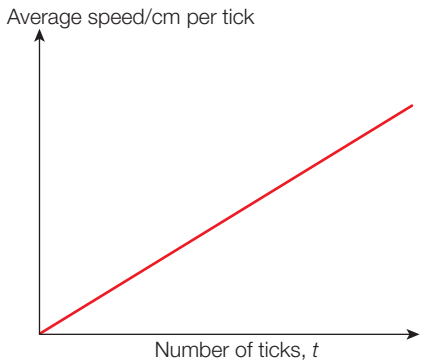


Figure 8.18

Results

The results are completed in table 8.7. If Newton's Second Law holds, the value of $m \times a$ will be constant, within the limits of experimental error.

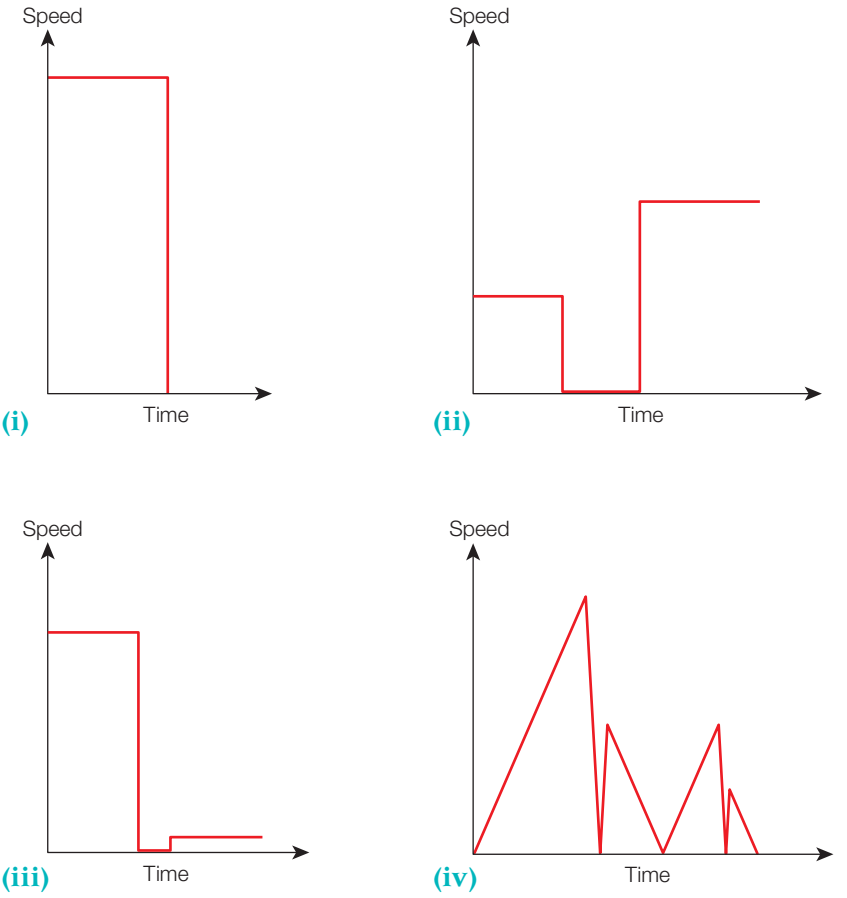
Chapter summary

- The rate at which distance is covered is called speed. Speed is a scalar quantity.
- Acceleration is the rate at which either speed or velocity changes with time.
- Acceleration could be either a scalar or a vector quantity: a scalar quantity if it is derived from speed, or a vector quantity if derived from velocity.
- The displacement of a body is the change of its position from its starting point. It has direction and is, therefore, a vector quantity.
- Velocity is the rate of change of displacement with time. It is a vector quantity.
- Speed–time and velocity–time graphs can be used to solve problems on straight-line motion.
- Newton formulated three laws that can be used to study motion:
  - Newton's First Law can be used to study motion under no external force.
  - Newton's Second Law is concerned with motion under an unbalanced force.
  - Newton's Third Law is concerned with the relationship between forces that come into play to cause momentum to change.
- Momentum is defined as mass × velocity. Its symbol is p and its S.I. unit is $kg\,m\,s^{-1}$.
- Newton's Third Law can be used to explain many common occurrences.

Answers to ITQs

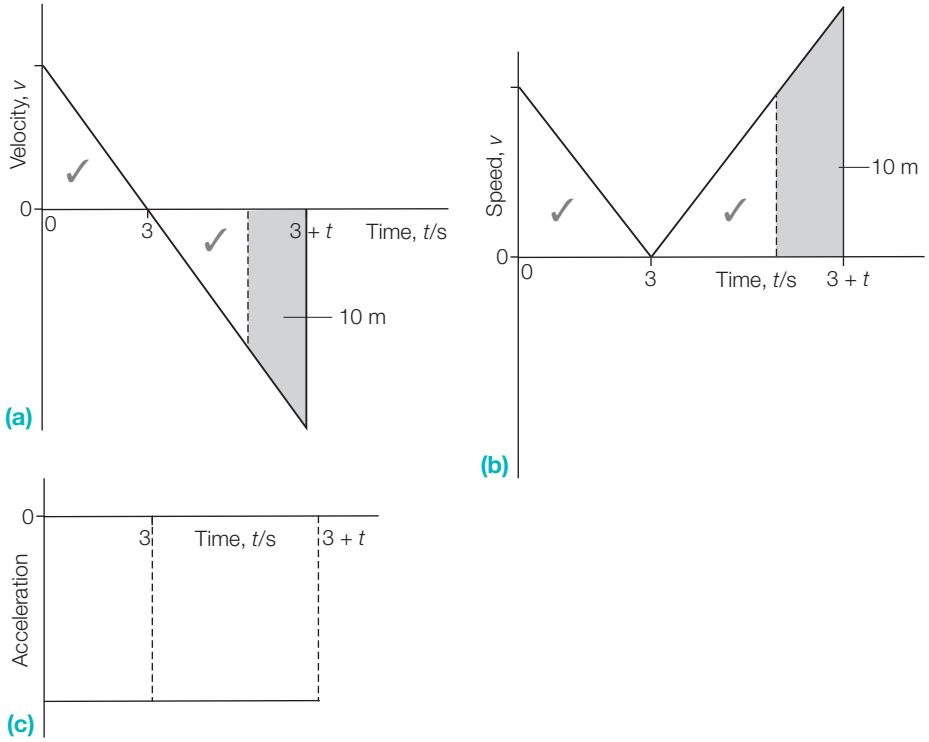
ITQ1 **(i)** 3 s; **(ii)** $0\,m\,s^{-2}$

ITQ2



ITQ3 $1.23\,m\,s^{-2}$

ITQ4



ITQ5 If we don't pedal the bike will slow down because of friction from the road at the axles and from the wind. A force therefore has to be used to overcome these resistive forces and bring the unbalanced force to zero.

ITQ6 10 m s^{-1} in the direction of the force.

ITQ7 The newton is that force which will give to a mass of 1 kg an acceleration of 1 m s^{-2}.

ITQ8 As the athlete moves lower on the trampoline, he encounters greater resistance from the extended springs. This resistive force slows the athlete down. This illustrates the second law: the external resistive force from the springs acts to reduce the momentum of the athlete. Third Law: on hitting the trampoline, the impact (shock) produced by the athlete on the trampoline on landing acts downward on the trampoline and the trampoline exerts an equal and upward shock on the athlete.

ITQ9 It is safer to use a soft trampoline because the springs, being softer, do not exert such a large retarding force the on the athlete and the athlete comes to rest more slowly. The 'jolt' felt by the ankles and knees is therefore less and so the likelihood of sustaining injury is thereby reduced.

Examination-style questions

(Where necessary, take the acceleration due to gravity to be 10 m s^{-2}.)

1 A passenger paces up and down in a stationary aircraft in a straight line at 0.50 m s^{-1}, covering a distance of 10 m at a time in each direction.

(i) Assuming that he paces up and down 5½ times before stopping, for his motion draw:

(a) a speed–time graph;

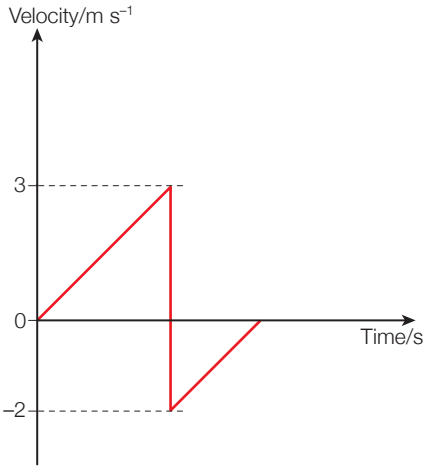
(b) a velocity–time graph;

(c) a distance–time graph;

(d) a displacement–time graph.

(ii) Calculate his average velocity over this period.

2 The velocity–time graph below represents the motion of a ball falling then rebounding in a vertical line.



(i) Find:
- (a) the time of fall of the ball;
- (b) the time of rise of the ball;
- (c) the total distance covered;
- (d) the overall displacement of the ball.

(ii) Describe the motion of the ball and sketch a displacement–time graph to represent this motion.

3 A motor-bike under test accelerates uniformly from rest and moves in a straight line. In 3 seconds it reaches a speed of 40 m s^{-1}. It then travels at this speed for 10 seconds, and then brakes to a stop in a further 2 seconds.

(i) (a) Draw a speed–time graph to represent the motion of the motor-bike.
- (b) Calculate the total distance the bike travels over the period of the trial.
- (c) Calculate the average speed of the bike over the period of the trial.

(ii) If the bike were to accelerate for a period and, straight after this, begin to brake in such a way that it braked over the same time as it accelerated, what would be the acceleration? Assume the test distance and duration are the same as in the first case.

4 A boy of mass 50 kg lets himself down out of a tree, falling from rest through a distance of 4 m. He brings himself to rest 0.5 second after his feet first touch the ground.

(i) Calculate:
- (a) his velocity on touching the ground;
- (b) his change in momentum in coming to rest;
- (c) the force exerted by the ground in bringing the boy to rest;
- (d) the boy's average deceleration as he is brought to rest.

(ii) Explain why, in falling from a great height, it is always a good idea to 'break your fall' either by choosing to fall on soft ground (if you can), or to take as long as possible to come to rest.

5 State and explain what is liable to happen when you:

(i) step off a chair that is standing on a smooth floor;

(ii) press against a wall with your palms, standing on a smooth floor in shoes with smooth soles;

(iii) step on a fresh mango skin that is lying on the pavement with its wet side uppermost.

6 Explain the following:
   (i) the action of moving a boat forwards by using a paddle;
   (ii) the use of a loop of rope around your ankles when climbing a tall, practically straight, tree (such as a coconut tree) without branches;
   (iii) why a humming bird can remain at rest by flapping its wings at a certain rate.

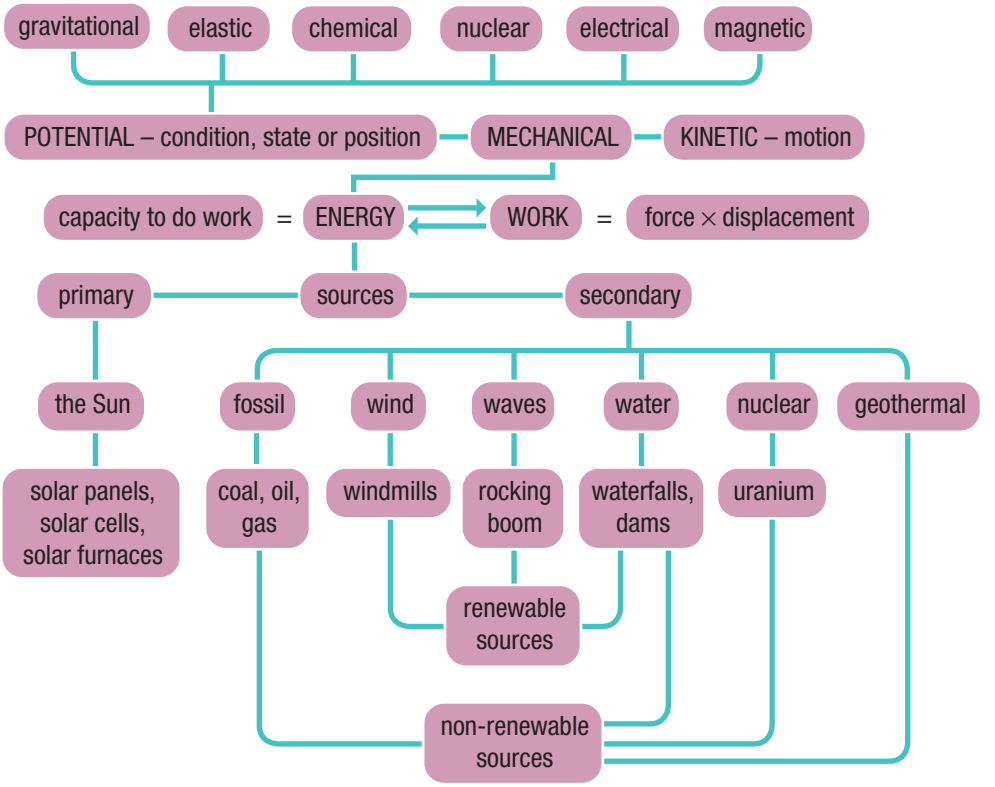
7 A steel ball of mass 0.50 kg falls from rest from a height of 2.0 m on to firm ground.
   (i) Find, by using a velocity–time graph:
      (a) how long it takes to reach the ground;
      (b) its velocity on touching the ground.
   (ii) On the ground, the ball comes to rest in 0.01 s. Calculate:
      (a) the deceleration of the ball;
      (b) how far the ball penetrates into the ground.
   (iii) Sketch a velocity–time graph for the entire motion of the ball.

8 When jumping on to the ground from a height, landing after parachuting, or skiing over bumpy ground, people are often told: 'Bend your knees or break your legs.' Use your knowledge of physics to explain why.

9 Energy, Work and Power

By the end of this chapter you should be able to:

- understand why energy is defined as the 'ability to do work' or the 'capacity for doing work'
- define work done by and work done against a force
- distinguish between potential energy and kinetic energy
- use the formula for the change of potential energy of an object to solve problems
- use the formula for the kinetic energy of an object to solve problems
- use the principle (or law) of conservation of energy to solve problems
- define power and use the definition to solve problems
- understand the meaning and the significance of the term 'efficiency'
- calculate efficiency in a given situation
- name some alternative sources of energy in the Caribbean and discuss the feasibility and importance of each of these sources as an alternative to conventional ones



Introduction

Most of us would agree that, if we didn't eat, we would soon become very frail and feel very weak. We would say that we lacked energy. Our muscles would

not be able to move our fingers to make them write, or our legs to make them run. This suggests that, in order to do things that we ordinarily do from day to day, energy is needed.

Why do we need energy? Let's think about what we do in writing and walking. We push a pen against the force of friction in writing, and we push our legs against friction on the ground in walking. So perhaps energy is needed in order to move against a force. This is one of the main purposes of having energy – to make it possible for something to overcome an opposing force and to move against it. This force may be mechanical, or it may even be electrical or magnetic. Whatever the type of force that has to be overcome, energy will be needed. Again, as long as an object moves in response to a force, we say that work is done either by or against that force. We seem to be saying, then, that there is a connection between energy and work.

We all know that the more work we do, the more energy we use up, and the more tired we feel at the end of it. It would be fair to say that we feel tired only because we have lost energy in doing that work. This leads to the thought that when work is done, energy is lost.

In this chapter we will discuss the important relationship between work and energy, and define and use the concept of 'work'. The concept of 'power' will also be covered. The important principle referred to as the 'conservation of energy' will be extensively used in worked examples, and we will discuss some of the sources of energy for domestic and industrial use in the Caribbean.

The meaning of energy

energy ›
work ›

Energy means the 'ability to do work' or the 'capacity for doing work', where **work** is said to be done when a force applied to an object causes that object to move. Since work and energy are directly related, and since a decrease in one results in an increase in the other, they have the same unit. The S.I. unit of energy, regardless of its type, is the **joule** (J).

joule ›

The energy in a system, considered as a general concept, may be placed basically into one or the other of two main categories, depending on the 'nature', or the circumstances which give rise to the energy. If the energy is due to the condition, position or state of the body, it is described as '**potential**'. If the energy derives from motion of the body it is described as '**kinetic**'. The adjectives 'potential' and 'kinetic' aptly describe these two basic 'natures' of energy. The word, 'potential', suggests 'can do' and the other word, 'kinetic', suggests 'in motion'. We may say, then, that whereas potential energy in a body enables that body 'to do' work by virtue of its condition, state or its position, kinetic energy in a body enables that body to do work by virtue of its motion (think of the destruction caused by hurricanes!). We must remember, however, that whatever the nature of an energy change, be it potential or kinetic, if the doing of work results, this work will always involve a force. Remember that there are many 'types' of force, e.g. elastic, magnetic, contact, gravitational, and so on (meaning that a force may be caused by many quite different circumstances).

potential energy ›
kinetic energy ›

CHAPTER 6

In chapter 6 we saw that a force may be gravitational, elastic, electrical, magnetic, electromagnetic, or nuclear. Each of these 'types' of force has a corresponding type of potential energy (or P.E.) associated with it. There are, therefore, as many types of potential energy as there are types of force.

Let us take a brief look now at the concept of 'work'.

Work

In physics, work is said to be done *by* a force when that force moves its 'point of application' through a distance *in the same direction as the force*. Work is a scalar quantity and the formula for work is:

Work done by (or against) a force = (value of the force) × (distance moved by the force in a direction parallel to the force)
or, $W = F \times d_{//}$

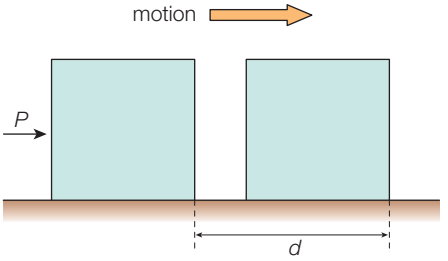


Figure 9.1

where the subscript '//' behind the d reminds the reader that the distance used must be parallel to the force.

In figure 9.1, the force P displaces the box through a distance d in a direction parallel to the direction of the force. The work done by the force in the direction of the force is therefore given by $W = P \times d_{//}$.

The unit of work is the same as that of energy, the joule (J). This is not surprising, because when two energies are subtracted, the resulting difference is work done.

One joule is defined as the work done by or against a force of one newton when that force moves its point of application through a distance of one metre in a direction parallel to the force.

This definition is, of course, based on the formula for work.

Worked example 9.1

A box is pushed by a force of 20 N over a distance of 4 m in a direction parallel to the force (figure 9.2). Calculate the work done by the force.

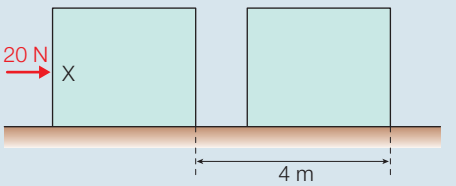


Figure 9.2

Solution

Here we say that work is done by the force, since the point of application of the force, X, moves in the same direction as the force does. The force and displacement are parallel. The work done is

W = force × displacement$_{//}$
= 20 N × 4 m
= 80 N m
= 80 J

Here N m = J because force and displacement are parallel (that is, in the same direction).

CHAPTER 5

Worked example 9.2

The same box as in Worked example 9.1 is now pulled by a rope inclined at 30° to the floor (figure 9.3). The tension in the rope is again 20 N. Calculate the work done by the force in this case if the box again moves through 4 m.

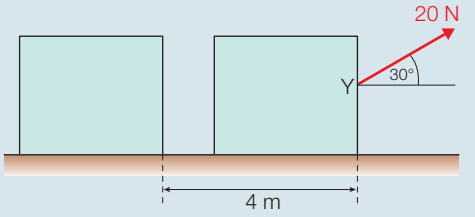


Figure 9.3

Remember that only vector quantities can be resolved, that is to say, replaced by two components each at right angles to the other. The displacement here can be resolved.

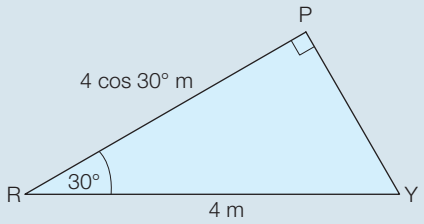


Figure 9.4

Solution

Using the definition given above, the work done is the force times the distance moved parallel to the force. From figure 9.4 this can be seen to be

W = force × distance moved parallel to the force

= 20 N × RP

= 20 N × 4 cos 30° m

= 69 N m

= 69 J

Again, 1 N m = 1 J when force and displacement are parallel.

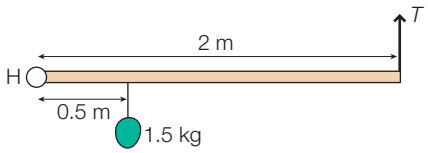


Figure 9.5

ITQ 1

In figure 9.5 the light rod is kept horizontal by a tension and a weight, as shown. The rod is hinged at H.

(i) Find the moment of the weight about the hinge, H, in S.I. units.

(ii) What is the expression for the moment of the tension about the hinge?

(iii) Why can we not use the unit 'joule' for the unit of moment?

(iv) Deduce the tension in the string.

Types of potential energy

There are six types of potential energy, namely gravitational P.E., elastic potential energy, electrical potential energy, magnetic potential energy, electromagnetic potential energy and nuclear potential energy. We shall be concerned only with the first two.

What is meant by 'gravitational P.E.'?

Remember that the adjective used to describe the energy depends on the type of force involved. If we consider a stone in the air above the Earth, we find that there is a gravitational force pulling the stone vertically downwards towards the Earth. We may say that the stone will be 'at ease' (literally!) when it is at rest on the ground and we will say that when the stone is 'at ease' on the ground its potential to do work is zero or its gravitational potential energy is, therefore, zero, since it cannot do any work (like breaking glass, for instance) when it is on the ground.

In order to break the glass, we must do the following:

1 Raise the stone above the ground to give to it the gravitational P.E. (G.P.E. for short) it needs to break glass.

2 Release the stone from the height to which it has been raised.

3 Allow the stone to fall in order to do work on the glass and so break the glass. (Do you know how the energy is used to smash the glass to pieces?)

Internet search term: what happens to atoms and molecules of things when they break?

We will consider the stages 1, 2 and 3 mentioned above in turn.

1 Raising the stone

In order to raise the stone:

- Someone's muscles must exert an upward force to support the stone by 'balancing out' the pull of gravity on it. This force, by Newton's first law, must be equal to the weight of the stone. So if the mass of the stone is m,

the weight is mg and so the force, F, the muscles exert to support the ball $= mg$.

- The hand exerting the force, F, must move vertically upwards through a distance, Δh, say, and this would entail work being done, since work is done when a force moves though a distance in a direction parallel to the line of action of the force.
- The formula for the work done is:
  (value of the force) × (distance moved in a direction parallel to the force)
  So the work done by the force F in this case is:
  $W = F \times \Delta h$
  If the force does work on the stone, then the stone must show that it has 'benefited', by acquiring more energy from my muscles.
- The energy acquired by the stone is gravitational potential energy, 'gravitational' because the force that had to be overcome was a gravitational force, and 'potential' because, by raising the stone higher the force has enabled it to do work on the glass when it strikes the glass. By raising the stone higher and then releasing it, the stone can break glass of greater thickness. The stone now has a greater ability (potential) to break the glass.
- So the additional G.P.E. of the stone in being raised through a height Δh,
  ΔE_p = work done by the force of my muscles = $F \times \Delta h$
  or $\Delta E_p = mg \times \Delta h$
  $= mg\Delta h$
  where mg = the weight of the body raised.

increase in G.P.E. of a body = (weight of the body) × (distance raised)
decrease in G.P.E. of a body = (weight of the body) × (distance lowered)

This is a formula you must know. You must also know how and when to use it.

Worked example 9.3

Using the formula: work done = force × distance moved parallel to the force, show that I do 1 joule of work when I raise 1 kg of mass through a vertical height of 10 cm.

Solution

Work done = (force my hand has to use to support the 1 kg mass)
× (vertical distance I raise the mass)
= (the weight of 1 kg mass) × (vertical distance mass is raised)
$= m \times g \times \Delta h$
$= 1\,\text{kg} \times 10\,\text{N kg}^{-1} \times 0.1\,\text{m}$ (units must all be S.I.)
$= 1\,\text{J}$ (also S.I.)

The 'unitary' definition of the joule is: the joule is the work done when a force of 1 N moves its point of application through 1 metre in a direction parallel to the force.

Worked example 9.4

A helicopter of total mass 4000 kg at a height of 500 m above the Earth loses power and falls to the ground. How much gravitational potential energy is released in the fall?

Solution

The vertical change in height of the helicopter = 500 m.

The formula to use is:

change in G.P.E. = (weight of body) × (vertical change in height)

Since the helicopter falls, it will lose G.P.E., so we have loss of G.P.E.,

$$\Delta E_p = mg \times \Delta h$$
$$= 4000\,\text{kg} \times 10\,\text{N kg}^{-1} \times 500\,\text{m}$$
$$= 20\,000\,000\,\text{J}$$

In order to avoid giving the answer with a string of zeros, we use the 'megajoule' as the unit of choice, and so

loss of G.P.E. = 20×10^6 joules

= 20 MJ (this is a lot of energy!)

We can imagine the destruction that such a crash can cause. Again we might ask: 'How is the energy of the crash used to cause such destruction?' See if the Internet can help you to find the answer.

ITQ2

Figure 9.6 shows two levels above the Earth's surface.

(i) How much work is needed to raise a mass of 0.5 kg from level 0 to level 1?

(ii) How much work is needed to raise a mass of 20 000 kg from level 0 to level 2?

(iii) How much energy is released when a 100 g mass falls from level 1 to level 0?

ITQ3

The gravitational field strength at the Moon's surface is only one-sixth of what it is at the Earth's surface.

(i) Where will a stone accelerate faster, at the Moon's surface or at the Earth's surface? How long will the stone take to fall from rest through 1 metre:

(ii) at the Moon's surface?

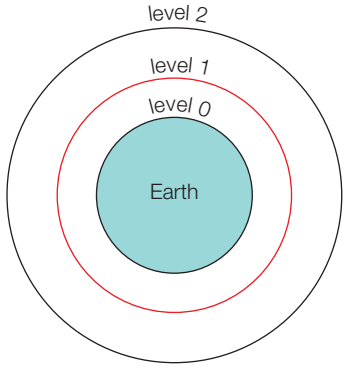
(iii) at the Earth's surface?

Take g to be 10 m s^{-2} on the Earth.

ITQ4

If I compress a spiral spring, will I be doing work against an elastic force? If so, will I be increasing the elastic potential energy of the spring? Give a reason for your answer.

Note that in the calculation we did not need to know the path taken by the helicopter as it fell to the ground. All we needed to know was the vertical distance through which the helicopter fell. We call this 'vertical distance through which the helicopter fell' the 'vertical displacement' of the helicopter. The helicopter fell from one geographical potential energy level above the Earth to a lower level, that of the Earth's surface.



level 0 = sea level
level 1 = 1000 m
level 2 = 2000 m

Figure 9.6

Other types of potential energy

A catapult pulled back (figure 9.7) can, if released, give energy of movement to a stone. A wound-up spring in a clock can deliver its potential energy to the clock's movement to make the hands of the clock go round. These two are both cases of elastic potential energy at work. Cases in which electrical potential energy, magnetic potential energy and nuclear potential energy are at work do exist, but are much less familiar in everyday life. Chemical potential energy is stored in our bodies, in coal, oil and natural gas, in the food we eat and in batteries. Potential energy, whatever its type, is energy that is stored in a body by virtue of the position of that body (as in gravitational potential energy) or by virtue of its condition (as in elastic potential energy or chemical potential energy).

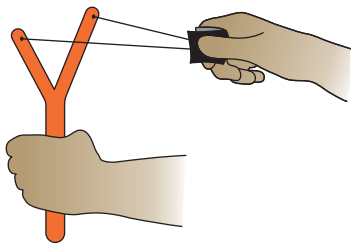


Figure 9.7

Kinetic energy

CHAPTER 12

Kinetic energy (K.E.) is 'motion energy'. It is energy which the entire body possesses and should not be confused with the energy of the molecules of the body. You will learn about the energy of molecules in chapter 12. It is the energy of the molecules which is responsible for temperature.

Kinetic energy is the energy possessed by a body by virtue of its motion. As a bicycle moves, its wheels move bodily (meaning from one place to another) and they also spin about their axles. The wheels will have kinetic energy by virtue of their centres of mass moving, and, also, by virtue of their rotation about their axles. Both of these types of motion contribute to the overall kinetic energy of the wheels, but we will not be concerned with kinetic energy due to rotation, only with the energy associated with the wheel as a whole moving from one place to another.

Caribbean countries are very familiar with the fury of hurricanes (figure 9.8). In 2013, hurricane Ernesto passed over several islands, leaving damage in its wake. The kinetic energy of the winds in this hurricane was so great that trees were uprooted, roofs were removed from many buildings, and galvanised roofing sheets were literally wrapped around tree-trunks. At sea, a storm involving high winds can cause similar damage to ships caught up in it. There, high winds and a raging sea both contribute to the energy of the storm.



Figure 9.8 The ravages of a hurricane – what effect have the high-velocity winds had on these trees?

Although the air is a 'light' substance, its tremendous speed in a storm provides it with sufficient energy to wreak havoc. In the case of waves, however, it is the mass of the water in the waves, more than their speed that is more responsible for the damage done. So kinetic energy depends on mass and on speed. As we shall see shortly, speed has the greater effect on kinetic energy. Doubling the speed increases the kinetic energy by four times, and tripling the speed increases kinetic energy by nine times. Doubling the mass of an object only doubles (not triples!) the kinetic energy. Clearly then, speed affects kinetic energy more than mass.

MATHEMATICS: algebra – direct variation

In fact the formula for the kinetic energy of an object is

kinetic energy of a body, $E_k = \frac{1}{2}$ (mass of the body) × (speed of the body)^2
or $E_k = \frac{1}{2}mv^2$

Whereas there are many different forms of potential energy, these depending on the type of forces involved, there is, fortunately, only one formula for kinetic energy.

The S.I. unit of kinetic energy, like that of all other forms of energy, is the joule.

ITQ5

Define the joule using the formula for kinetic energy.

Worked example 9.5

A missile of mass 2 kg is projected vertically upwards and rises 100 m into the air. Calculate its initial kinetic energy at the time of projection

Solution

The steps in the method are:

1 Use the velocity–time graph to find an equation for the time, t, the missile takes to rise 100 m.

2 Use the value obtained for t to find the initial velocity the stone must have in order to have zero velocity after t seconds. Remember that the deceleration of the missile will be 10 m s^{-2}.

ITQ6

Use the steps given immediately above to calculate the value of:

(i) the initial velocity, and

(ii) the initial kinetic energy of a cricket ball of mass 160 g which is projected vertically upwards to a height of 20 m.

ITQ7

Which of the two catches described below of a cricket ball will cause a greater 'sting' when the ball is caught:

(a) a slip catch at first slip of a ball travelling at 25 m s^{-1}; or

(b) a 'skied' ball falling vertically from a height of 25 m?

Using these steps leads to (velocity)^2 = 2000 m^2 s^{-2}.

So from the formula for K.E., we have

$$\text{initial K.E.} = \tfrac{1}{2}mv^2$$
$$= 0.5 \times 2 \times 2000\,\text{J}$$
$$= 2000\,\text{J}$$

Have you noticed that K.E. is proportional to (speed)^2? This means that if the speed of a hurricane wind doubles, the damage it will do will be four times as great. It means also that if the driver of a motor car doubles his/her speed and wants to make an emergency stop, the car's stopping distance will increase by four times – if the driver uses the same braking force, of course!

The principle (or law) of conservation of energy

inter-conversion of forms of energy ›

An important fact of life is that energy is constantly changing from one form to another both in nature and in our everyday lives; we call this change of the form of energy from one form to another 'transformation of energy' or the **inter-conversion of forms of energy**. This is true for cases involving both human and animal activity and it is also true in the case of plants and other living organisms. In fact, it is true of all nature and not only for living things.

A good example of a natural occurrence, the effects of which we are very aware of, are, again, hurricanes, perhaps the most familiar natural disaster in the Caribbean. They occur only in the summer months when there is a plentiful supply of heat energy in and over the ocean. Here this heat gives rise to increased movement of air, and this eventually leads to the steady moment of large bodies of air spiralling at high speed. When this air 'makes landfall' the kinetic energy in it is dissipated (used up) in causing destruction of property and other disastrous effects. As the hurricane progresses leaving more and more devastation in its wake, it loses kinetic energy and slows down. An

energy flow diagram ›

energy flow diagram for these changes might look as follows:

heat in and above the ocean → kinetic energy of the air → energy involved in destroying property

As a simpler and more familiar example of the law of conservation of energy being demonstrated, take the case of a child's swing. I use the chemical P.E. stored in my arms to raise the child and the seat of the swing along an arc, thereby doing work against the weight of the child and the seat. When I release the seat, both child and seat fall to a lower level, thereby losing G.P.E. From the conservation law just discussed, the G.P.E. lost must change into a different form of energy which clearly must be kinetic or 'motion energy', since the child and the seat will now be moving faster. In fact the lower the seat gets, the faster the seat moves, showing that as the swing loses G.P.E. it gains kinetic energy. This is because of the conservation law. As one form of energy falls, another form rises. We have assumed that there is no frictional resistance or drag acting on the swing to slow it down and cause it to lose kinetic energy.

CHAPTER 6

The fact that energy is continuously changing its form means that energy is not really being created that was not there before. What is there now was always there but was present in a different form. This fact is expressed briefly by a law (or a principle) called the 'law of conservation of energy', which says that:

Energy cannot be created nor destroyed; if energy disappears in one form, it re-appears in another.

primary source of energy ›

Of course we must remember that what energy is available to us on Earth has come (and is still coming) from the Sun, which we regard as the '**primary**' source of all forms of energy that exist.

There are many other cases where we use our own strength to increase the potential energy of a system, such as the following examples:

(a) pulling the bow before releasing an arrow;
(b) winding a mechanical clock;
(c) raising the seat of a child's swing before releasing it;
(d) a wrestler climbing on to the ropes of the wrestling ring before jumping from his perch on to his opponent.

The energy needed to do each of the things mentioned comes from the muscles in the body and these muscles would have been provided with energy from food. We would write the energy flow diagrams for these examples as follows:

(a) chemical P.E. in muscles of the arm → work done in pulling the bow → elastic potential energy in the bow

(b) chemical P.E. in muscles of the arm → work done in winding up the spring → E.P.E. in the clock-spring

(c) chemical P.E. in muscles of the arm → work done in raising the seat → G.P.E. in swing

(d) chemical P.E. in muscles of wrestler's arms and legs → work done in climbing on to the ropes → G.P.E. of wrestler on the ropes

ITQ8

Write an energy flow diagram for the energy change(s) which take(s) place when each of the following occurs: (i) a light appears when a torch is switched on; (ii) you ride a bike uphill; (iii) an electric kettle boils; (iv) a lorry accelerates; (v) a lorry brakes to a standstill.

What happens to the energy used to push the switch of the torch in ITQ8? Can you write an energy flow diagram for the changes here?

Other cases that occur in nature are:

- water at the top of a waterfall;
- ripe fruits (like coconuts) falling from trees.

Of course we are all aware that we have to eat in order to keep alive and our cars and other motor vehicles need to be 'fed' with fuel in order to serve their purpose and this is always concerned with providing energy of one sort or another. It is clear that in these cases the foods we eat provide chemical potential energy for our bodies and in the course of performing our daily activities; this energy is transformed (converted) into other forms some of which are shown below:

- work in overcoming forces in climbing, lifting, etc.;
- moving various parts of the body like limbs or eyes;
- using the senses like seeing or thinking;
- producing heat to keep the body at the correct temperature.

Much the same might be said of the fuel used by vehicles. The chemical potential energy which the petrol or diesel contains is converted to other forms of energy in the vehicle, such as motion energy (called kinetic energy) or heat, some of which is produced by moving parts having to overcome friction, but mostly produced as a result of the chemical reactions taking place in the engine.

Frictional force and energy

Friction always opposes motion. If, in trying to field a cricket ball, I skidded along the ground, I would soon stop moving. If I push a book along a table-top it may not fall off the edge. If someone tries to move a car when the driving

wheels are resting on slippery ground, they may well fail, since the driving wheels may just spin round and round. Sometimes friction can be to our advantage and sometimes to our disadvantage.

We know from the law of conservation of energy that, as the speed of a moving object falls and its kinetic energy falls, this is accompanied by the appearance of another form of energy. If the force opposing the motion is a **frictional force** then the form of energy that replaces the kinetic energy lost is **thermal energy**, more familiarly known as heat. There are countless examples of this. Rubbing your hands together, sharpening tools on a grindstone, lighting a match, a spacecraft re-entering the Earth's atmosphere, and motor vehicle braking are some familiar examples.

frictional force ›
thermal energy ›

Heat is thermal energy that passes from a hotter body to a colder body because of their difference in temperature.

In all of these cases, whether the friction is between two solid surfaces (solid–solid) or between a solid surface and a fluid (solid–fluid, called drag), thermal energy (which we will call 'heat', for the time being) is produced just the same. When frictional forces of any kind bring about a fall in kinetic energy, heat is always produced to replace the kinetic energy lost.

Now that we have covered all the relevant issues involved in getting the stone to fall from a height and break the glass below, let us do a recap. We go back to the stone that was raised to gain the energy needed (or potential energy) to break the glass when it falls. On raising the stone:

(i) chemical energy in my arm muscles → work done to raise the stone

(ii) work done to raise the stone → gain of G.P.E. of the stone

2 The stone is released

As it fell, the stone lost G.P.E. and gained K.E. The energy flow diagram for this is:

(iii) G.P. E. in the stone → K.E. in the stone

3 The stone hits the glass

When the stone came to rest on the glass, it lost all its K.E. Work was done on the glass to destroy it, and:

(iv) K.E. in the stone → work done to smash the glass

Find out from the internet how energy lost by the stone is used to fracture the glass.

It is clear, then, that chemical energy in the muscles of my arm 'ended up' being transferred to the glass to break it. This is a good example of the transformation (or inter-conversion) of energy from one form to another.

Power

power ›

A weight-lifter may be more 'powerfully built' than a 100 m runner, but it is doubtful whether, in a 100 m race, he can show his '**power**' better than the less powerfully built sprinter. This suggests that 'power' in physics and 'powerful' in the ordinary sense of the word are not necessarily related. Whereas 'powerful' ordinarily means 'looking muscular and strong', in physics 'power' means 'the rate at which energy is transformed' or 'the rate at which work is done'. The runner may therefore be able to do work or to convert energy at a faster rate than the weight-lifter and so have a greater power (in the scientific sense) than the weight-lifter.

In physics, power is defined as follows:

$$\text{power} = \frac{\text{work done}}{\text{time taken to do the work}}$$

so

$$\text{power} = \frac{\text{force applied} \times \text{displacement}_{//}}{\text{time taken}}$$
$$= \text{force applied} \times \frac{\text{displacement}_{//}}{\text{time taken}}$$
$$= \text{force applied} \times \text{velocity} \quad \text{(since displacement} \div \text{time} = \text{velocity)}$$

This shows that another important formula for power developed is:

$$P = F \times v$$

We have divided 'work done' (a scalar quantity) by 'time' (another scalar quantity) to obtain power. it follows that the result is a scalar. So 'power' is a scalar quantity.

About power

If the power is required at any moment, then

power at that moment = force × velocity (at that moment)

However, if average power is required (over a period), then

$$\text{average power} = \frac{\text{total work done}}{\text{total time taken}}$$

or

$$\text{average power} = \frac{\text{total energy transferred}}{\text{time taken to do this}} = \frac{W}{t}$$

Because $\text{power} = \frac{\text{work done}}{\text{time taken}}$

then work done = power × time taken

or $W = P \times t$

Worked example 9.6

I can raise a bucket of cement mix of mass 12 kg through a vertical height of 8 m in 10 seconds (figure 9.9). Calculate the average power used in raising the bucket against the gravitational force.

Solution

The total work done to increase the gravitational potential energy of the cement mix is equal to the gravitational force times the vertical distance the bucket rises:

$$W = mg \times h$$
$$= 8\,\text{kg} \times 10\,\text{N kg}^{-1} \times 8\,\text{m}$$
$$= 640\,\text{J}$$

Thus

$$\text{power used (average over a period)} = \frac{\text{work done}}{\text{time taken}} = \frac{640\,\text{J}}{10\,\text{s}} = 64\,\text{J s}^{-1}$$

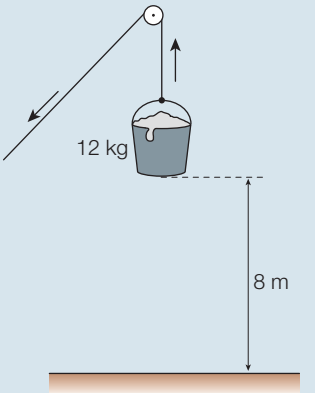


Figure 9.9

Note!

In Worked example 9.6, we have considered only the work done in raising the cement mix against the force of gravity. We have assumed that the speed of the bucket remained constant throughout the motion and, therefore, that there was no increase in the kinetic energy of the bucket. If there has been no increase in velocity, there has been no acceleration, and so there has been no net force acting on the bucket. This means that the force used has been only that required to overcome the gravitational pull. All the work done went into increasing the G.P.E. of the bucket.

The unit of power

watt ›

Power is defined as the rate of transfer of energy, or the rate at which work is done (which is the same thing). The unit of power is the **watt** (W), named after James Watt, a British engineer who applied scientific principles to the development of the steam engine.

> A rate of transfer of energy of 1 joule per second is defined as one watt, $1\,W = 1\,J\,s^{-1}$.
> One kilowatt = 1 kW = 1000 W.

We could therefore say that the average power used in Worked example 9.6 is 64 watts (64 W).

IT:Q9

A weight-lifter can lift 'weights' of mass 300 kg through a vertical height of 1.8 m in 4 seconds. Calculate the average power he develops in lifting the weight.

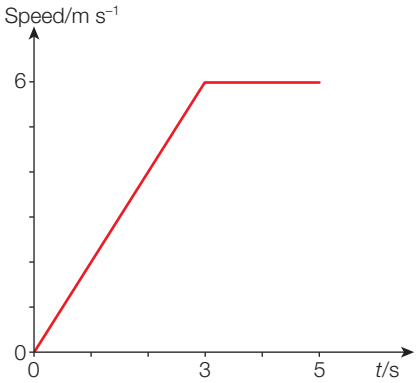


Figure 9.10

Worked example 9.7

The graph in figure 9.10 shows a part of a short sprint done by an athlete. Her mass is 50 kg. She accelerates uniformly for 3 seconds to a speed of 6 m s^{-1} and then maintains this speed for 2 seconds longer. Calculate (assuming no wind resistance):

(i) the power she develops in attaining the speed of $6\,m\,s^{-1}$;

(ii) the average power developed over the 5 seconds.

(Ignore the work the athlete does in overcoming frictional resistances.)

Solution

Method 1

(i) We first find the acceleration and the accelerating force that produces that acceleration:

$$\text{acceleration} = \frac{\text{change in speed}}{\text{time taken}} = \frac{(6-0)\,m\,s^{-1}}{3\,s} = 2\,m\,s^{-2}$$

$$\text{force producing the acceleration} = \text{mass} \times \text{acceleration} = 50\,kg \times 2\,m\,s^{-2} = 100\,N$$

We now use the formula

$$\text{power} = \text{force applied} \times \text{velocity} = 100\,N \times 6\,m\,s^{-1} = 600\,W$$

This is the power she attains on reaching the speed of $6\,m\,s^{-1}$.

(ii) The power needed to maintain this speed is (the force acting while the speed is constant) × (the constant velocity). The force (due to drag), as assumed before, is negligible and so the power needed to maintain the speed of $6\,m\,s^{-2}$ is small and will be ignored. So

$$\text{average power} = \frac{\text{total work done in reaching the speed of } 6\,m\,s^{-1}}{\text{total time taken for the short sprint}}$$

$$P = \frac{F \times d}{t} = \frac{100\,N \times (\frac{1}{2} \times 3\,s \times 6\,m\,s^{-1})}{3\,s} = 300\,W$$

This is the average, since her power will vary with her speed, assuming that her net force remains constant.

Note, interestingly, that the power used is proportional to the speed while the net force remains constant. It therefore increased as the speed increased.

Method 2

The average power = $\frac{\text{initial power + final power}}{2} = \frac{0 + 600\,\text{W}}{2} = 300\,\text{W}$, as before.

Method 3

Rather than finding the work done by the net (accelerating) force in attaining the speed of $6\,\text{m s}^{-1}$, we could find the increase in the kinetic energy of the athlete (energy transformed) and divide that by the time taken. This gives

$$\text{increase in KE} = \tfrac{1}{2} \times 50 \times 6^2 - \tfrac{1}{2} \times 50 \times 0^2$$
$$= 900\,\text{J}$$

So

$$\text{average power developed} = \frac{900\,\text{J}}{3\,\text{s}}$$
$$= 300\,\text{J s}^{-1} = 300\,\text{W} \quad \text{again as before.}$$

The answers to Worked example 9.7 suggest that, while the body is accelerating, the power it must develop increases to a large value, since not only must it do work to gain kinetic energy, but it must also overcome frictional resistance. Once the body reaches its cruising speed (top speed), it only has to overcome frictional resistances (though small) to maintain that speed.

A car just starting off in first gear needs a large pulling force to overcome its inertia (see chapter 6) and get it moving. The power here should be as large as possible so that the force at this time could be as large as possible for a low speed (car just starting to move). As the car accelerates and the speed rises, this pulling force gets smaller, if we assume that the power of the engine remains constant. The car is now moving faster in a higher gear and the pull of the engine is smaller than it was in first gear.

CHAPTER 6

ITQ10

I can run up a flight of stairs 3 m high (not long!) in 5 s. What is the power I generate if my mass is 65 kg?

Efficiency

Efficiency is usually associated with devices that use energy in order to fulfil a purpose and tells us how well that energy is used to complete the job. There are many such devices in use. Some devices help us to transfer energy of one kind into energy of another in order to perform the task. Others may not convert the energy supplied into another form, but may make the task to be performed easier to perform. We call such devices 'machines'. A simple example is a hammer. Since we cannot pull a nail out of a piece of wood with our bare hands, we use a hammer. How does the hammer work to do this?

We first apply a force to the hammer and as we pull backwards on the hammer handle, we do work (force moves through a distance) and transfer energy to the hammer. What the hammer is expected to do now is to transfer as much of this energy as possible to the nail (through its 'claw') in order to draw the nail out of the wood against the frictional resistance (called the 'load') of the wood and what fraction of this energy is transferred by the hammer depends on the efficiency of the hammer. The force that is applied is called the effort. The size of the effort needed, in its turn, depends on the efficiency of the hammer. So how is efficiency determined?

We would ideally like to find that all the energy we put into the hammer was used to get the job done, that is, to draw the nail. If we were so fortunate as to be able to achieve this, we would say that the process was 100% efficient,

meaning that all the energy applied to the hammer was properly used to draw the nail and none of it was wasted. But we find in practice that we can never achieve such an ideal result, since there is no device in which some of the input energy is not lost in the course of using it through having to overcome friction (that cannot be avoided, since surfaces will move against others, involving frictional forces) or for some other undesirable, but necessary, reason.

There is no device, whether natural or man-made, that is so perfect, as to have an efficiency of 1. They all fall short of perfection in that only a fraction of the energy supplied to the device to perform the specific task is usefully used.

The energy lost through friction (which we have discussed earlier) appears as heat and is deemed to be **'wasted' energy**. This generation of heat as a device is used is very typical of mechanical machines which employ mechanical forces (see chapter 6). Electrical and electronic machines are not exempt from energy loss in the form of heat either since, apart from the fact that they may have moving parts (think of the fan, the hard-drive and the CD-drive of your computer which rotate about their axles) and this will result in frictional forces being present and therefore heat being generated. Currents flowing in conductors will invariably also produce heat.

'wasted' energy ›

CHAPTER 6

In effect, then, how much of the energy input to a device is usefully converted by the device to perform the expected task (or 'to do the job'), e.g. raise a load (as in a fork-lift), give a light (as in a torch), produce sound (as in a loudspeaker), light up your computer screen and provide information (as in your computer), etc., is expressed as a percentage, the formula used being:

The Greek letter 'ε' is pronounced 'ep-s(eye)-lon' or 'ep-si-lon'.

$$\text{efficiency of the device} = \frac{\text{energy used to do the job}}{\text{total energy provided to the device}} \times 100\ (\%)$$

$$\text{or, efficiency, } \varepsilon = \frac{\text{energy usefully converted}}{\text{total energy provided}} \times 100\ (\%)$$

The efficiency may also be given in terms of the power input to the machine and the power output from it. The input power is the rate at which energy is supplied to the machine to get the job done, and the output power is the rate at which the machine converts energy to do the job. It is clear that if the numerator and the denominator of the expression above are each divided by the time the machine is used, we will have

$$\text{efficiency, } \varepsilon = \frac{\text{output power}}{\text{input power}} \times 100\ (\%)$$

We can also use the law of energy conservation to describe efficiency. Since energy is conserved, then energy put into the machine = energy obtained from it to 'do the job' + energy used otherwise (or wasted energy)

or, using symbols, we can say that

$$E_{in} = E_{out} + E_{wasted}$$

where E_{wasted} is the total 'wasted' energy. Dividing throughout by E_{in}, we have

$$1 = (E_{out}/E_{in}) + (E_{wasted}/E_{in})$$

But E_{out}/E_{in} = the efficiency, ε, of the machine, so from the last equation,

$$1 = \varepsilon + f_w,$$

f_w being the fraction of the input energy that is wasted.

We have then $\varepsilon = 1 - f_w$.

This shows then that the efficiency of a machine, ε, is not constant, but depends on the fraction of the input energy that is wasted. It tends to become smaller and smaller as the load (which is the magnitude of the job to be done) becomes greater.

We would all agree that you would not use a fork-lift truck to lift a carton of soft-drink. Why?

ITQ 11

A small motor draws 120 J of electrical energy from the mains to lift a book of mass 1 kg through a vertical distance of 8.0 m. It takes 5 s to do this. Calculate: (a) the power input to the motor; (b) the work done by the motor in lifting the book; (c) the fraction of the input power that is wasted.

Worked example 9.8

A warehouse fork-lift takes 4 s to lift a crate of mass 1000 kg through 5 m working at a rate of 20 kW. Five per cent (5%) of the wasted energy in the forklift is used to lift moving parts of the fork-lift and the rest is lost as electrical heating and frictional heating. Calculate:

(i) the efficiency of the fork-lift at this load;

(ii) the value of the total energy wasted;

(iii) the amount of heat developed in lifting the crate.

Solution

To say that the fork-lift is working at the rate of 20 kW is to say that the power drawn from the source (normally batteries) is 20 000 W.

The useful energy converted by the fork-lift is that energy which raises the crate and this energy = $mg\Delta h$ = 1000 kg × 10 N kg^{-1} × 5 m = 50 000 J.

(i) Efficiency = $\frac{\text{useful energy converted to work in lifting the crate}}{\text{total electrical energy supplied to the fork-lift}} \times 100\%$

$= \frac{50\,000\text{ J}}{20\,000\text{ W} \times 4\text{ s}}$ (since energy = power × time)

$= 62.5\%$

(ii) If 62.5% of the energy supplied to the fork-lift is usefully used (to lift the load), then 37.5% is wasted.

Wasted energy = 37.5% of 80 000 J

= 30 000 J

(iii) If 5% of this wasted energy is used to lift certain moving parts, then 95% of it is wasted as heat.

So, amount of energy lost as heat = 95% of 30 000 J

$= (\frac{95}{100}) \times 30\,000\text{ J}$

= 28 500 J

MATHEMATICS: percentages

Worked example 9.9

A certain make 'X' of low energy electric lamp is advertised as being six times as efficient as a certain brand 'Y' of incandescent bulb.

(i) What do you understand by the statement above?
The amount of light obtained from an incandescent lamp bulb for every watt of power it receives from the source is as low as 8%.

(ii) Express the statement above in another way. What is the form of energy of the remaining 92% of the energy converted in the incandescent bulb? Write an energy flow diagram for the energy conversion in the bulb. What happens to the energy produced by the bulb eventually?

(iii) How many joules of light energy are obtained from the low energy lamp when it takes 100 W of electrical energy from the mains?

Solution

(i) The statement means that: if both lamps are supplied with the same amount of electrical energy, the low energy lamp will convert six times as much of this energy to light as the incandescent lamp will.

(ii) Another way of saying this is to say that the efficiency of the incandescent lamp is 8%. The remaining 92% is converted to heat:

electrical energy from the mains → (heat + light)

The heat produced is transferred to the atmosphere, where it is lost irretrievably.

Carry out your own research to determine whether the following claim made by a certain manufacturer of low energy lamps is a reasonable one:

'The low energy lamps we produce have an efficiency which is about 6 times that of most incandescent bulbs.'

(Remember that the purpose of the lamp is to produce light and not heat!)

Energy sources

The Sun is our primary source of energy

primary energy source ›

For millions of years the Sun has been our major source of energy and it will continue to be so for a long time to come. Our existence depends on the Sun, which we call the **primary source** (from *primus,* the Latin word meaning 'first') of our energy. Just consider what would happen if the Sun stopped shining! There would be no plant life and we would all starve, for there would be no crops. There would also be no rainfall, since rain clouds are formed from water vapour in the atmosphere and this can only take place if there is sufficient warmth to produce evaporation from the seas and the oceans. Not only would there be no human beings, there would be no animals either. It seems, then, that no form of life would exist if there were no sunshine. It would also be much too cold for both animals and plants to survive.

The Sun is called the primary source of our energy because it is the source from which most of the secondary forms of energy (like winds, water in waterfalls and waves) come.

While the Sun will continue to shine for a very, very long time to come, the most widely used sources, namely fossil fuels like coal, oil and natural gas, will not. Supplies of these resources will come to an end eventually, if not soon. It makes good sense, therefore, for all countries, including Caribbean ones, to look elsewhere for other sources of energy. It would, indeed, make better sense if these alternative sources were renewable, this meaning that they will 'renew' themselves and not be lost to us once used, like nuclear power, for instance. Possible sources of this kind are wind and the tides in the ocean, hydropower from waterfalls and, of course, the Sun. Attempts have been made in the last few decades by a handful of islands to use solar power, but this has been only on a domestic scale. Attempts have also been made in a few islands to develop geothermal resources, but again, on a very limited scale.

So how real is the possibility that Caribbean countries will be able to develop some of the alternative sources like solar energy, wind energy, hydro-electric energy and the like more generally?

What about wood, charcoal and biomass?

These are important sources of energy in the context of developing counties like those in the Caribbean. Wood and charcoal are still much used in rural areas of some Caribbean territories, where they are easily available (e.g. Guyana and Belize with extensive forested areas). Since it is possible to replace trees that have been used for fuel and there is always a plentiful supply of vegetable waste, all three of these sources can be considered to be renewable.

biomass ›

The use of **biomass** has not, however, really taken root in the Caribbean, although official attempts were made in the 1970s to generate interest in the idea as a means of supplying energy relatively cheaply for domestic use.

Prospects for the development of alternative sources

It does appear that there has not been a serious attempt to investigate and develop on a national scale many of the alternative resources available to Caribbean countries, and all the territories depend on the use of fossil fuel of one kind or another. Alternative sources like nuclear fuels and hydroelectric power are not generally within their reach, either because of initial costs or because the natural resources (waterfalls, for example) do not exist. So what are the alternatives available to these countries?



Figure 9.11 Electric car.

The car in figure 9.11 uses an electric motor for much of its journey. The petrol engine cuts in only when needed. The car is intended to be charged overnight from the mains, but could alternative energy sources be used to provide the mains power that is needed for charging?

There is a lot of tidal movement in the seas around the Caribbean and many of our islands are hilly and exposed to strong and steady winds. Might there be the possibility of investing in tidal power generators, windmills (figure 9.12) and wind farms? There certainly seems to be a possibility in these areas although, because of the high probability of hurricanes in the region, these could be either risky or expensive ventures to undertake. With the exception of Guyana, Caribbean territories do not have sufficiently high or wide waterfalls with a reliable flow of sufficiently large volumes of water to make the building of dams or the harnessing of waterfalls a worthwhile possibility. It is very likely, however, that Guyana, with its generous supply of falls of all sizes will soon have the first hydropower generator among Caribbean countries when work on the Amaila Falls Hydro Project in the west of the country is completed.



Figure 9.12 Windmill.



Figure 9.13 Solar panel.

In all of the Caribbean territories there is abundant sunshine, and because of this, a few islands have made use of this resource and have set up solar panels (figure 9.13), but for domestic purposes only. Again, as with wind turbines, this is regarded as a risky venture to invest in on a larger scale because of the prevalence of hurricanes in the region.

Is there a likelihood that sea waves could someday be harnessed to provide power stations with the steam power they need to drive turbines? Might all the territories in the Caribbean be able to develop their biomass resources to the extent where these can be used to provide sufficient energy at least for domestic purposes? Is it possible to develop 'gasohol', in those countries where sugar cane is cultivated, for use in motor vehicles? Can '**bagasse**', produced from the ground sugar cane, be more widely used as fuel to reduce the strain on conventional fuels? Will it someday be possible to drive solar-powered battery-operated motor cars? These are probably some of the directions in which countries in the Caribbean might look for a partial solution to the energy 'crisis' that looms ahead.

'Gasohol' is alcohol fuel produced by fermentation of sugar from sugar cane.

bagasse ›

Bagasse is what remains after the sugar cane has been crushed and the sugar juice extracted. These remains are dried and used as fuel in some sugar factories in the Caribbean.

Chapter summary

- Energy is a measure of the 'capacity' or the 'degree of ability' of a body to do work. The unit of energy is the joule.
- Energy is a scalar quantity.
- Work done on or by a body is the product: (force acting on that body) × (displacement of the body in the direction of the force).
- If work is done on a body, there is a change of energy in that body.

- Mechanical energy is of two main types, potential and kinetic.
- Potential energy is energy due to position or energy due to condition or state.
- Kinetic energy is energy due to motion.
- Potential energy is of different forms: gravitational, elastic, electrostatic, magnetic and nuclear.
- Power is the rate of doing work or the rate at which energy is converted or transferred. It is a scalar quantity.
- The unit of power is the watt. One watt is one joule per second.
- When energy changes take place as a result of work done, they do so in accordance with the principle (or law) of conservation of energy.
- This law states that energy cannot be created nor destroyed. If there is a change (like a decrease) of energy in a system or a body in one form, it will be accompanied by a corresponding change (an increase) in some other form.
- In nature and also in our homes, offices and factories, energy changes are taking place all the time. These changes are referred to as the inter-conversion of energy.
- The sources of energy are the Sun (called the primary source) together with others (secondary sources), among which are some that depend on the presence of the Sun.
- Those sources which are natural processes that renew themselves and are constantly present are called 'renewable sources'.
- Those sources which do not renew themselves and which, once used, cannot be renewed in a short time are called non-renewable sources.
- Supplies of non-renewable sources, especially coal, oil and natural gas, are being used up at an increasing rate. It is necessary, therefore, for us to use energy with care (conserve it) and to look for and develop alternative sources of energy that may be available to us in the Caribbean.

Answers to ITQs

ITQ1 **(i)** 7.5 N m; **(ii)** $2T$; **(iii)** this is a moment and not a work; **(iv)** $T = 3.75$ N

ITQ2 **(i)** 5000 J; **(ii)** 400 MJ; **(iii)** 1000 J

ITQ3 **(i)** On the Earth's surface **(ii)** 1.1 s; **(iii)** 0.45 s

ITQ4 Increasing, since I will be doing work to compress the spring and so elastic potential energy will be gained by the spring.

ITQ5 One joule is the kinetic energy that a mass of 2 kg possesses when moving with a speed of 1 m s^{-1}.

ITQ6 32 J

ITQ7 The slip catch.

ITQ8 **(i)** electrical energy in battery → heat and light in bulb
(ii) chemical energy in leg muscles → G.P.E. in cyclist and bike
(iii) thermal energy in heat source → latent heat in boiling water
(iv) chemical energy in petrol → G.P.E. in lorry
(v) kinetic energy in lorry → thermal energy in braking system

ITQ9 1350 W

ITQ10 390 W

ITQ11 **(i)** 24 W; **(ii)** 80 J; **(iii)** 0.33

Examination-style questions

1 Fill in the blanks in the following passage with the appropriate words from the list supplied. Each word may be used once, more than once, or not at all.

work **force** **done** **application** **capacity** **parallel**

direction **same** **motion** **net potential** **kinetic**

The energy of a body is a measure of the ..... of that body to do ..... Work is done by a ..... when that ..... moves its point of ..... through a distance that is ..... to the force. If a net force acts on a body in the .... as that in which the body moves, that body will accelerate in the direction of the ..... This means that the body will gain ..... energy. The ..... energy gained is equal to the .... by the ..... force.

2 (i) In which of the two cases stated below will the kinetic energy lost by the body be greater when the body hits a hard surface and is brought to rest almost instantly?

(a) An iron ball of mass 50 g that falls through a vertical height of 50 m.

(b) The same iron ball travelling with a speed of 90 km h^{-1}.

(ii) What is the difference between the two energy losses?

3 A boat of mass 2000 kg accelerates from rest to a speed of 9.0 m s^{-1} in 10 s. The total frictional resistance to the motion of the boat is 1500 N. Calculate:

(i) the acceleration of the boat;

(ii) the net force that produces this acceleration;

(iii) the work done by the engine of the boat in producing this acceleration;

(iv) the work done in overcoming the frictional resistance;

(v) the average power developed by the engine;

(vi) the power at which the engine was working at the end of 10 seconds.

4 A moving stair-case raises 40 passengers through a vertical height of 10 m in 30 s. The average mass of the passengers is 60 kg. Calculate:

(i) the gravitational potential energy gained by the passengers;

(ii) the power required to do the above.

(iii) If, instead of a staircase, a lift was used to raise the same number of passengers in the same time, do you think the power required would be greater than, less than or equal to this value? Give a reason for your answer.

5 A bullet of mass 10 g enters a target at 100 m s^{-1} and leaves it 0.01 s later at 40 m s^{-1}. Calculate:

(i) the rate of change of momentum of the bullet;

(ii) the resistance offered by the target;

(iii) the loss of kinetic energy of the bullet;

(iv) the power used by the target in slowing down the bullet;

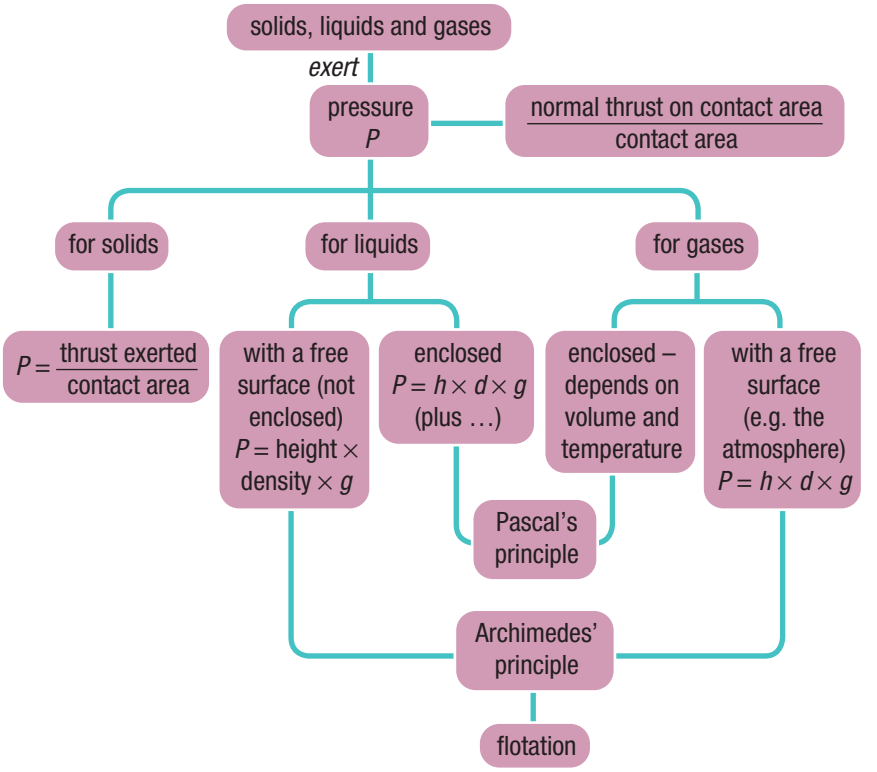
(v) the deceleration of the bullet;

(vi) the distance travelled by the bullet in passing through the target.

10 Pressure and Archimedes' Principle

By the end of this chapter you should be able to:

- define the pressure exerted by a solid, or a liquid on a surface
- apply this definition to calculate the pressure on a surface
- define the S.I. unit of pressure, the pascal, and convert a pressure expressed in pascals into another unit
- explain how pressure is produced by solids and liquids
- relate the pressure at a point in a fluid to its depth and the density
- apply your understanding of the concept of 'pressure' to explain some common situations in everyday life
- understand and apply Pascal's principle to explain the behaviour of common hydraulic systems
- use the 'law' that pressures are the same at a given horizontal level in a liquid that is stationary and continuous
- use the principles of fluid pressure to deduce Archimedes' principle
- use Archimedes' principle to obtain the 'principle of flotation' and then predict whether a body would float or sink when placed in a given fluid
- perform calculations using Archimedes' principle and the principle of flotation



Introduction

The idea of force is familiar to all of us but we are not so familiar with the concept of pressure. We know that a net force acting on a solid body will cause that body to move. It is not so obvious, however, that a difference in pressure could cause a liquid or a gas to move. This is what happens when we turn on our water taps to get water and our gas jets to obtain gas to light a flame. When we drink from a straw, we use a pressure difference to get the liquid up the straw. Many of the incidents that take place during a hurricane in which household property is lost or destroyed can be put down to differences in air pressure inside and outside buildings. These are only a few of many instances of the consequences of pressure difference that we meet from day to day.



Pressure

pressure ›

Pressure may be exerted on a surface by a solid, by a liquid or by a gas, which applies a thrust to that surface. The pressure on the surface is that quantity which expresses the normal (perpendicular) thrust per unit area of contact between the agency (i.e. the solid, liquid or gas) and the surface. If the thrust is applied over a small area the pressure will be large; if it is applied over a larger area, the pressure will be smaller in value.

The word 'agency' is used here to mean the 'thing' that is causing the thrust, i.e. the solid or liquid

If I stand in my shoes on soft, muddy ground, I run the risk of sinking into the mud. But if I stand on a piece of board and distribute my weight over the area of contact between the board and the ground, then I am less likely to sink. In the second case I would be distributing my weight over a larger area (that of the board), and so the perpendicular thrust over unit area of the ground would be smaller than before, when the area of contact was only that of my soles. In other words, the pressure when I stand on the board would be smaller.

Pressure depends on the force (thrust) acting perpendicular to the surface that experiences the pressure and the area over which this perpendicular force is distributed.

The definition of pressure

What is meant by pressure? We define the pressure on a surface as follows:

$$\text{pressure on surface} = \frac{\text{thrust acting normal (perpendicular) to surface}}{\text{area of contact between agency and surface}}$$

or pressure, $P = \frac{F}{A}$

normal ›

The force must be measured **normal** (perpendicular) to the surface and the area is the area of contact between the agency and the surface. This formula can be used to find the pressure exerted by solids and liquids.

The S.I. unit of pressure

Since

$$\text{pressure} = \frac{\text{normal force}}{\text{area of contact}}$$

the S.I. unit of pressure is given by

$$\text{S.I. unit of pressure} = \frac{\text{S.I. unit of force}}{\text{S.I. unit of area}} = \frac{\text{newton}}{\text{metre}^2} = \text{N m}^{-2}$$

pascal ›

This unit, N m^{-2}, has been named the '**pascal**' (Pa) after Blaise Pascal, a French theologian and mathematician of the 17th century. Other units of pressure, based on the metric system, like N cm^{-2} and N mm^{-2}, are also used in physics, when they are more appropriate. Still other units of pressure, like the torr and the bar (not named after scientists) are also found in branches of science.

See whether you can find definitions of the torr and the bar on the internet and identify in which branch of science each one is used.

The abbreviation 'Pa' for 'pascal' has a capital 'P', because we are using the name of a person as the unit. We do not just use the first letter 'P' because there is another unit that has the abbreviation P for a different physical quantity.

Although 'pressure' is calculated from 'force', a vector quantity, pressure is nevertheless a scalar quantity.

Pressure due to solids

The pressure exerted by solids on a surface is always due to a thrust exerted by the solid on that surface. If a solid block rests on a surface, the pressure exerted by it on the surface is due to the thrust exerted on that surface (as a result of the pull of gravity) at right angles to the surface. If I exert a force on a surface with a solid (like a walking stick), however, the pressure on the surface will not be due to the weight of the stick, but to the force transmitted to the surface by means of the stick. An example of the first case is given below.

Worked example 10.1

Figure 10.1 shows a wooden beam, of mass 400 kg, length 20 m and cross-section 20 cm × 20 cm. Calculate the pressure exerted on a horizontal floor when it is:

(a) placed on one of its smallest faces (in $N\,cm^{-2}$);

(b) laid on one of its other faces (in $N\,m^{-2}$).

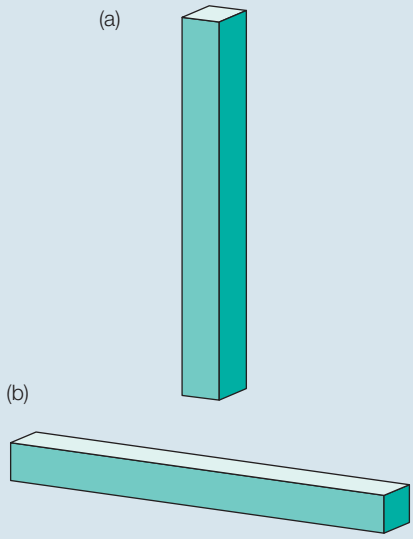


Figure 10.1

Solution

(a) The pressure on one of its smallest faces will be:

$$P = \frac{\text{downward thrust}}{\text{area of contact}}$$
$$= \frac{4000\,N}{20\,cm \times 20\,cm}$$
$$= \frac{4000\,N}{400\,cm^2}$$
$$= 10\,N\,cm^{-2}$$

as in figure 10.1 (a).

To say that the pressure on the ground is $10\,N\,cm^{-2}$ is to say that each square centimetre of area of the ground in contact with the beam experiences a downward force (or thrust) of 10 N.

To express this result in pascals, we work as follows:

$1\,cm^2$ experiences a force of 10 N

and $1\,m^2 = 10^2\,cm \times 10^2\,cm = 10^4\,cm^2$

Therefore, $1\,m^2$ would experience a force of 10 N (for $1\,cm^2$) $\times\ 10^4 = 10^5$ N.

Therefore, the pressure will be $10^5\,N\,m^{-2}$ or 10^5 Pa.

So a pressure of $10\,N\,cm^{-2}$ is equivalent to a pressure of 10^5 Pa.

Another way of approaching the matter is to say that the pressure,

$$P = 10\,\text{N cm}^{-2} = \frac{10\,\text{N}}{(1\,\text{cm})^2} = \frac{10\,\text{N}}{(10^{-2}\,\text{m})^2}$$
$$= 10\,\text{N} \times (10^4\,\text{m}^{-2})$$
$$= 10 \times 10^4\,\text{N m}^{-2}$$
$$= 10^5\,\text{N m}^{-2}$$
$$= 10^5\,\text{Pa}$$

MATHEMATICS: mensuration; conversion of units; indices

(b) If the same beam were laid on its side (figure 10.1 (b)), however, the pressure would be (downward thrust)/(new area of contact with the ground)

The pascal is an S.I. unit, so we need to use the S.I. unit of length, the metre, in calculating the area of contact. This area = 20 m × 20 cm

$$= 20\,\text{m} \times 0.20\,\text{m}$$
$$= 4\,\text{m}^2$$

so the new pressure,

$$P = \frac{\text{downward thrust}}{\text{area of contact}}$$
$$= \frac{4000\,\text{N}}{4\,\text{m}^2}$$
$$= 1000\,\text{N m}^{-2}$$
$$= 10^3\,\text{Pa} \qquad \text{(or 1 kPa)}$$

MATHEMATICS: algebraic functions

Which is 100 times smaller than before. So, by making the area of contact 100 times larger, we have made the pressure 100 times smaller.

This is an inverse relationship.

Since the thrust will remain the same for both positions of the beam on the ground, if the area of contact is changed, the pressure will vary inversely as the area of contact, since

$$P \propto \frac{1}{A}$$

So P varies inversely with A.

Worked example 10.2

Express the pascal in base units.

Solution

$$1\text{ pascal} = \frac{1\text{ newton}}{1\text{ metre}^2}$$
$$= \frac{1\,\text{kg m s}^{-2}}{1\,\text{m}^2}$$
$$= 1\,\text{kg m}^{-1}\,\text{s}^{-2}$$

Worked example 10.3

When I stand on both feet I exert a pressure of 2200 Pa on horizontal ground. How much pressure will I exert if I stand on:

(i) one foot;

(ii) tiptoes on both feet and the contact area with the ground on one foot reduces to one-third of what it was on one foot?

Solution

(a) The pressure when I stand on both feet = 2200 Pa.

By standing on one foot only, I am reducing the contact area of my foot with the ground by a factor of ½.

Since the pressure varies inversely with the area of contact ($P \propto \frac{1}{A}$), then reducing A by ½ will cause P to change by a factor of the inverse of ½, that is by $\frac{1}{0.5}$ or by 2.

So the new pressure on one foot will be 2 × the old pressure, that is, 2 × 2200 Pa = 4400 Pa.

(ii) If I stand on tiptoes on one foot, the contact area reduces by ⅓, and so the new pressure will be $\frac{1}{(1/3)}$ or 3 times as large as before.

The new pressure on tiptoes on one foot will be 3 × 4400 Pa or 13 200 Pa.

By standing on tiptoes on both feet, I am doubling the area of contact and, because of the inverse relationship between pressure and area of contact, the pressure will change by a factor of the inverse of 2, which is ½.

The new pressure on tiptoes on both feet will be ½ of what it was on one foot, that is, ½ of 13 200 Pa, or 6600 Pa.

This is easier to understand using a table:

Area of contact	Change introduced – old area multiplied by	Old pressure multiplied by	Pressure/Pa
Two feet: 2*A*	–	–	2 200
One foot: *A*	$\times \frac{1}{2}$	$1 \div \frac{1}{2} = 2$	4 400
One foot on tiptoes	$\times \frac{1}{3}$	$1 \div \frac{1}{3} = 3$	13 200
Two feet on tiptoes	× 2	$1 \div 2 = \frac{1}{2}$	6 600

ITQ 1

What will be the new pressure on the ground:

(a) if I put one of these beams on top of another similar one;

(b) if I stack the beams, four wide by three high? (Hint: Use variation.)

MATHEMATICS: direct and inverse variation

Pressure due to liquids at rest

hydrostatic pressure

Pressure due to liquids at rest is called **hydrostatic pressure**.

In the last section we saw that the pressure due to a solid is due to the weight of that solid or, more generally, the thrust exerted by that solid on the surface. Since liquids also have weight, we may use the same formula to calculate the pressure due to a liquid.

Unconfined liquids in open vessels

Consider a vertical column of liquid in a large cooking or storage vessel, as shown in figure 10.2. At the base of the column, we can find the pressure, P, as follows:

$$\text{pressure due to liquid} = \frac{\text{weight of column of liquid}}{\text{contact area between liquid and base of vessel}}$$

$$= \frac{(\text{mass of liquid column}) \times g}{(\text{area of base of liquid column})}$$

$$= \frac{(\text{volume} \times \text{density}) \text{ of liquid} \times g}{\text{area of base of liquid column}}$$

$$= \frac{(\text{area of base} \times \text{height} \times \text{density}) \text{ of liquid} \times g}{\text{area of base of liquid column}}$$

$$= \frac{(A \times h) \times \rho \times g}{A} \quad \text{(using symbols)}$$

And so $P = h \times p \times g$ (cancelling A)

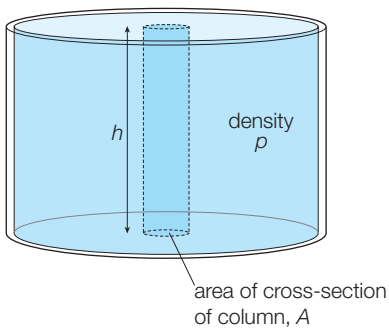


Figure 10.2

The symbol Δ (the Greek capital letter 'delta') is used to mean 'the change in …'. Therefore, because ρ and g are constants, for a change in depth, Δh, the change in pressure, ΔP is given by

$\Delta P = \Delta h \times \rho \times g$

The formula $P = h \times \rho \times g$ applies at any point in a stationary liquid which is at a depth h below the surface. The pressure at a point on a surface is the same for any inclination of that surface and it always acts at right angles to that surface. It acts and has the same value in all directions at a particular point (figure 10.3). So regardless of the orientation of the surface, the pressure at a point on it is given by

P = (depth of point in the liquid) × (density of liquid) × g

A particular feature of liquid pressure is that it acts in all directions at a point in the liquid – unlike solid pressure.

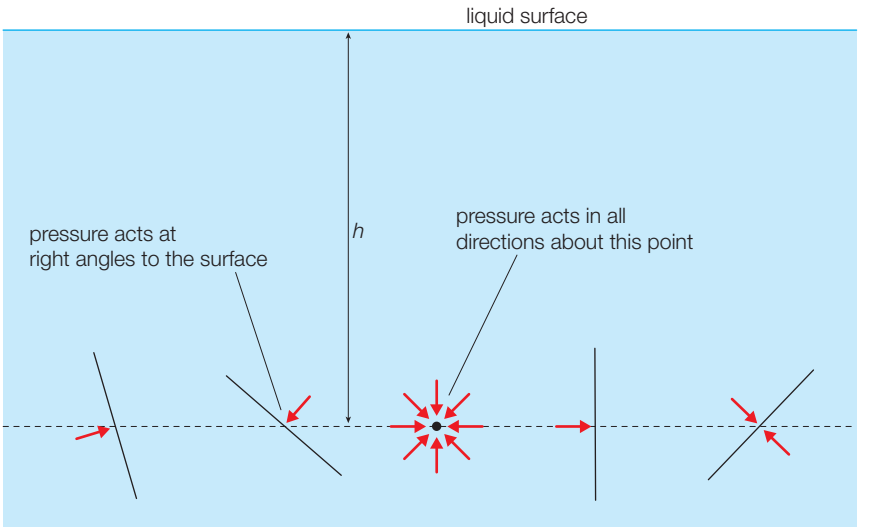


Figure 10.3 The pressure at all these points at depth h is the same.

Worked example 10.4

Calculate the hydrostatic pressure exerted on the hull of a sunken ship at a point X that is 1000 m below the surface of the water (figure 10.4). Take the density of sea-water to be 1012 kg m^{-3}.

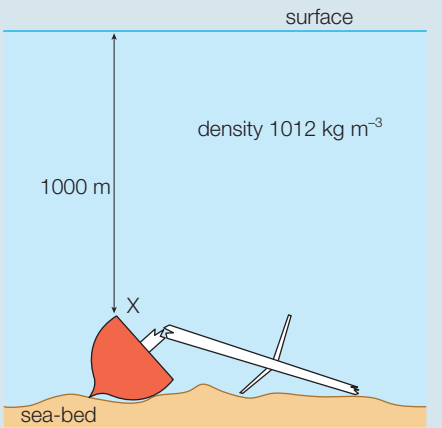


Figure 10.4

Solution

Regardless of the orientation of the surface where the point X occurs, the pressure at that point is given by

$P = h \times \rho \times g$

where h = the depth of the point below the surface.

Substituting, we have

$P = 1000\,\text{m} \times 1012\,\text{kg m}^{-3} \times 10\,\text{N kg}^{-1}$

$= 1.012 \times 10^{7}\,\text{N m}^{-2}$

$= 10.12 \times 10^{6}\,\text{Pa}$

$= 10.1\,\text{MPa}$

MPa = megapascals; 1 MPa = 10^6 Pa.

Practical activity 10.1

Aim

To show that the pressure exerted by a liquid depends on the depth of the liquid

You will need:

- a metal or plastic vessel about 40 cm deep with holes (about 1 mm in diameter) bored along a line down the side at intervals of about 5 cm
- a cork to fit the top of the cylinder tightly.

Method

1. Fill the cylinder with water to the brim and quickly insert the cork into the cylinder.
2. Observe what happens at the holes along the side of the cylinder.

Observation

Water spurts out of the holes in the cylinder, each stream with a different throw.

The 'throw' of the water that leaves the holes differs with the depth of the hole below the water level in the cylinder – the lower the level of the hole the greater the throw of the water.

Inference

The pressure at the positions of the holes varies with the position of the individual hole – the lower the hole the greater the pressure, and the greater the pressure the greater the 'throw'.

Conclusion

The pressure acting at a point where a hole is situated depends on the depth of the point – the lower the hole the greater the pressure.

This demonstration will be a qualitative one in which we show the 'qualitative relationship' between the depth of the hole and the pressure at the hole. A more exact demonstration is described later. Before going on to that activity, however, we will discuss a little later a very useful device for measuring liquid pressure called a manometer.

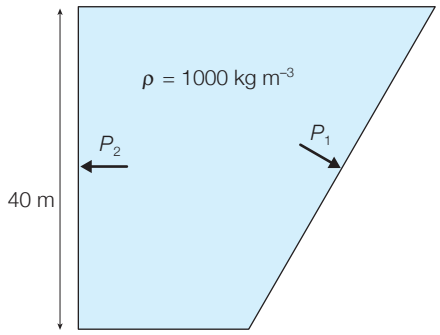


Figure 10.5

Worked example 10.5

The cross-section of a canal is shown in figure 10.5. The depth of the canal is 40 m and the density of the water it contains is 1000 kg m^{-3}. The sloping side makes an angle of 60° with the horizontal.

(i) Calculate the pressure, P_1, acting at a point half-way down the sloping side.

(ii) Calculate the pressure, P_2, acting at a point half-way down the vertical side.

(iii) Show the direction of the pressure in each case.

Solution

(i) The angle made by the sloping side with the vertical is of no importance when considering the liquid pressure on it. All that matters is the depth below the surface of the point where the pressure is required. This depth is 20 m. So, the pressure, P_1, acting at a point half-way down the sloping side is

$$P_1 = h \times \rho \times g$$
$$= 20\,\text{m} \times 1000\,\text{kg m}^{-3} \times 10\,\text{N kg}^{-1}$$
$$= 200\,000\,\text{Pa}$$
$$P_1 = 200\,\text{kPa} \text{ (since } 1\,\text{kPa} = 1000\,\text{Pa)}$$

(ii) Since the depth of this point is the same as that of the first, the pressure, P_2, there will have the same value, i.e. $P_2 = 200\,\text{kPa}$.

(iii) The pressure must always be perpendicular to the surface in question. The directions of the pressures will therefore be as shown by the arrows in the diagram of figure 10.5.

Liquids connected to each other will find their own level

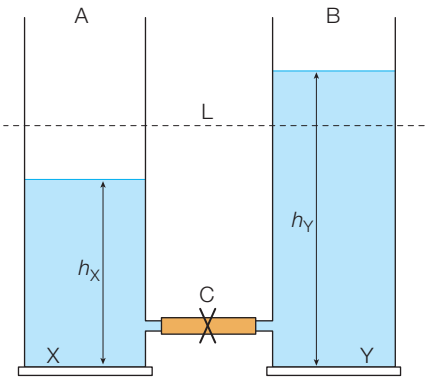


Figure 10.6

Figure 10.6 shows two quantities of the same liquid in separate vessels, A and B, on a table, joined by a tube carrying a closed clip, C. The surfaces of the two liquids are at different heights above the table.

As long as the liquids are kept separate by the closed clip, C, their levels remain unchanged. If the clip is opened, their levels will change, the higher one falling and the lower one rising. Why does this happen?

In figure 10.6, the liquid pressure at point X in vessel A is less than that at point Y in vessel B, because the depth of X in A is less than that of Y in B. Because of this pressure difference, the liquid in the joining tube will move in the direction Y to X from a point where the pressure is higher to another point where the pressure is lower. This is by no means a strange occurrence, since we witness the same principle when we turn on our garden hose. As this happens, the level of liquid in B will fall and that in A will rise. When there is no longer a pressure difference, the final level becomes L, and the two liquids have equal depths.

What would happen if the vessels were standing on tables at different levels and the liquid levels were:

(i) different at the start?

(ii) the same at the start?

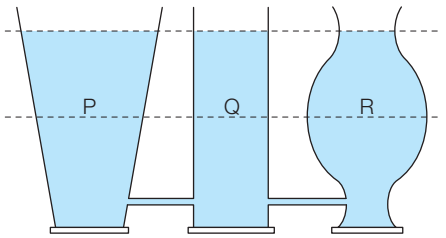


Figure 10.7 The surfaces of liquids that are connected together will be at the same level. The pressures at P, Q and R are also equal.

The vessels containing the liquids could be of any shape, since liquid pressure does not depend on the shape of the vessel containing the liquid, but only on the depth of the point at which the pressure is taken. The depths of liquid in the vessels shown in figure 10.7 will therefore all have to be the same if the bases are all on the same horizontal level. It follows, too, from this 'law' that all points on the surface of a liquid at rest will be in the same horizontal plane.

Figure 10.8 shows two vessels containing the same liquid. The heights of the liquid are different, and the levels of the vessels are themselves different. The vessels are connected by a flexible tube provided with a clip, which is closed. When the clip is opened, the two liquid surfaces will move toward a common level.

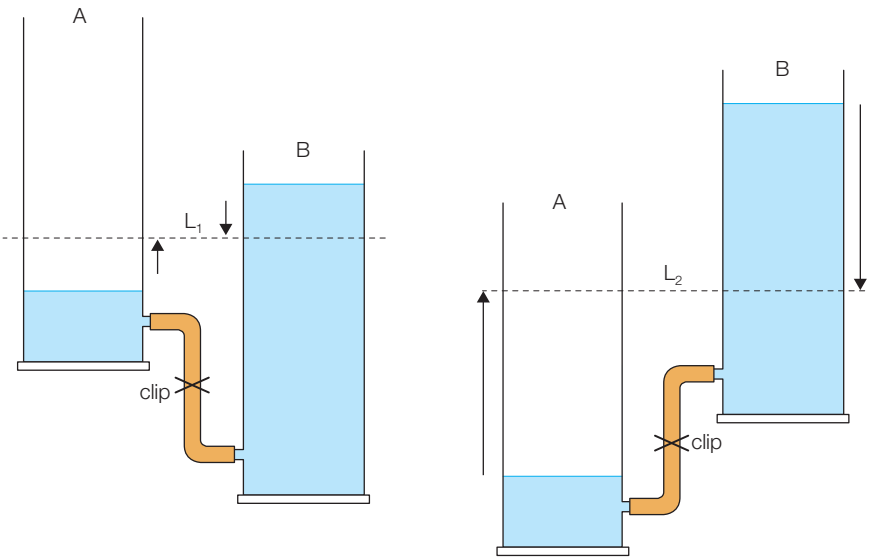


Figure 10.8 Connected liquids will find a common level.

Unconfined liquids and Pascal's law

Pascal's law ›

A liquid in a closed vessel behaves differently from an unconfined liquid. This difference is summed up in **Pascal's law** (Pascal's principle), which states that:

> **Pressure applied to any point of a fluid in a closed vessel is transmitted equally to every other point in the fluid.**

This principle forms the basis of many hydraulic systems, for example the braking system of cars, and hydraulic jacks for raising motor vehicles. The principle can be demonstrated in the laboratory by the following simple experiment.

Practical activity 10.2

Investigating Pascal's law

You will need:

- thin (non-rigid) plastic bottle
- cork
- sewing needle.

Method

1. Make a number of very fine holes all over the plastic bottle with the sewing needle.
2. Fill the bottle with water to within about 2 cm of its mouth.
3. Insert a tightly fitting cork into the mouth of the bottle.
4. Push the cork further into the bottle to increase the pressure in the bottle.

Note: an alternative to pushing the cork further into the bottle is to squeeze the bottle.

Observation

The water spurts out of the bottle through the fine holes scattered over the bottle. The throw of the thin streams is much the same from all the holes.

Inference

1. There is now a greater pressure acting at the positions of the holes than there was before. The pressure must have increased as a result of pressing the cork further into the bottle (or squeezing the bottle).
2. This increased pressure was present at all the holes in the bottle, since the water was seen to leave the bottle at each of the holes.
3. The increased pressure was much the same at all the holes.

Conclusion

Because (i) water spurts out of the bottle, and (ii) through all the holes, then (i) there must be an increased pressure in the bottle, and (ii) the increased pressure produced by pushing the cork into the bottle is transmitted to all points on the inner surface of the bottle. Since the throw of the streams was much the same for all, then the pressure transmitted was much the same at all the holes.

A confined liquid also behaves differently from a solid when the pressure at any point on it is increased. In our experiment in Practical activity 10.2, the increased pressure was 'felt' at all points in the liquid, including those where the liquid and the container met. If the bottle contained a solid or was entirely solid glass, this would not be the case. If a glass bottle is squeezed around the middle, the pressure would be 'felt' by the bottle only where the fingers touched the bottle. If the glass bottle contained liquid, the liquid would experience no change in pressure. This means, then, that a solid will not transmit a pressure; it will, however, transmit a force.

When the pressure in an enclosed fluid is increased, the increase in pressure is transmitted throughout the fluid. This pressure increase, if applied to a piston of large area, can give rise to a very large force, F. So a small force, f, applied to a piston of small area (why should this area be small?) can produce a much larger force F. Here is an example.

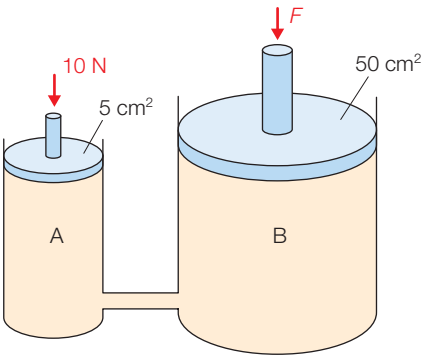


Figure 10.9

Worked example 10.6

Figure 10.9 shows two cylinders, A and B, each full of the same liquid and connected to each other by a tube. The piston in A has an area of 5 cm² and that in B has an area of 50 cm². A force of 10 N is applied to the piston in A. What force is needed on the piston in B to keep the pistons stationary?

Solution

We know that

$$\text{pressure} = \frac{\text{force}}{\text{area}}$$

so

$$\text{pressure under piston in A} = \frac{10\,\text{N}}{5\,\text{cm}^2}$$

$$= 2.0\,\text{N cm}^{-2}$$

For the pistons not to move, this must be the transmitted pressure to be applied downward on the piston in B to prevent it from moving. If the force applied to the piston in B as a result of this transmitted pressure is F, then from the formula $F = P \times A$, the force created under the piston in B,

$$F = 2.0\,\text{N cm}^{-2} \times 50\,\text{cm}^2$$

$$= 100\,\text{N}$$

To prevent the piston in B from rising under the action of the upward force resulting from the transmitted pressure, we must apply an equal downward force to the piston in B. This force is therefore 100 N.

ITQ2

If the radii of the pistons A and B were respectively 1 cm and 5 cm, and a force of 10 N was applied to piston of cylinder A, what would be the resulting upward force on piston B?

The manometer

manometer ›

Now that we have discussed Pascal's principle, we can discuss how the **manometer** works for the measurement of pressures that are a little larger or a little smaller than atmospheric. A diagram of this is shown in figure 10.10.

The manometer is essentially a simple U-tube containing water, oil or (sometimes) mercury. It is the best device to use if the pressures to be dealt with are small. If the pressure to be measured is just below or just above atmospheric, then either water or an oil can be used in the U-tube (figure 10.10 (a)). But for pressures above this value and up to about 1½ times atmospheric, it will be more convenient to use mercury in the manometer (figure 10.10 (b)).

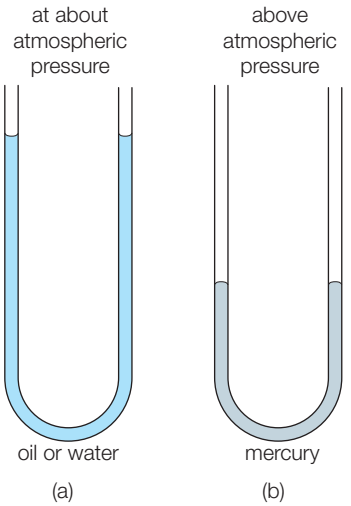


Figure 10.10

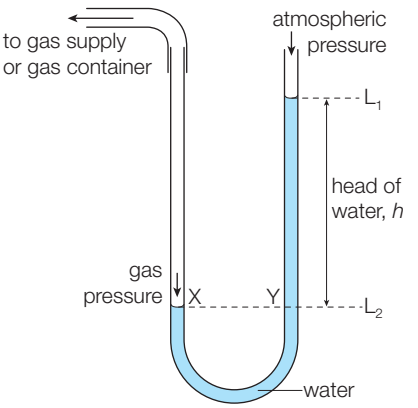


Figure 10.11

How does the manometer work?

If you blow down one limb of the manometer, the liquid level in that limb falls while the other one rises (see figure 10.11). The difference between the liquid levels is used to find the actual pressure being measured. For example, in the figure where the manometer is connected to a laboratory gas supply, the pressure of the supply is acting on the lower meniscus on the left, labelled X. By Pascal's law, since the liquid is continuous and at rest, the pressure at the level L_1 in the diagram is the same in both manometer limbs. Thus if the liquid meniscus on the left is denoted by X and that at the same level on the right is Y, we can say that

pressure (due to the gas) at X, P_{gas} = pressure at Y (inside the tube) due to the column h of liquid on the right above the point Y + the pressure of the atmosphere above, P_{atmos}

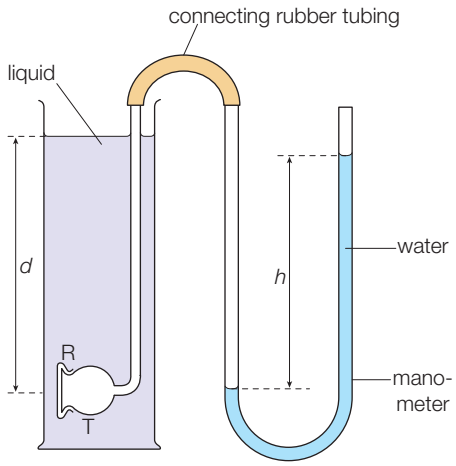
or $P_{gas} = h\rho g + P_{atmos}$

head of liquid ›

Since in practice we are more interested to know by how much the gas pressure is greater than the pressure of the atmosphere, rather than the actual pressure of the gas, we take the difference between the two.

This difference ($P_{gas} - P_{atmos}$) is what is of interest and that is what is most often measured. The length, h, of liquid separating the levels of the two menisci is called the head of liquid. The hydrostatic pressure due to this head of liquid is $h\rho g$. So when we use a manometer we often measure only the head of water, oil or mercury, depending on which liquid is used in the manometer. We would say that the gas pressure is h cm of water or h cm of oil, or the pressure (if high) is h cm of mercury.

Did you know that blood pressure is measured in mm of mercury above atmospheric pressure? What doctors measure is how much our blood pressure is above the pressure of the atmosphere. So if someone's systolic pressure is 130, it means that the pressure exerted by the person's heart to pump blood into the arteries is 130 mm of mercury above atmospheric pressure.



T = thistle funnel
R = thin rubber stretched over mouth of funnel and kept securely in place with rubber band

Figure 10.12

ITQ3
Calculate in pascals the pressure exerted by a column of mercury 130 mm high. The R.D. of mercury = 13.6.

ITQ4
Explain how you would use the apparatus shown in figure 10.12 to demonstrate that pressure acts in all directions about a point in a liquid.

ITQ5
Why is it more convenient to use mercury in a manometer for measuring high pressures?

Practical activity 10.3

Testing whether liquid pressure varies directly as liquid depth

Aim

To investigate how the pressure at a point in a stationary liquid varies with the depth of the point below the surface of the liquid

You will need:

- a 1000 ml measuring cylinder graduated
- a thistle funnel with a sheet of thin 'cellophane' (or 'cling film' or thin rubber) stretched over the mouth of the funnel and held securely in place with rubber bands (the 'seal' between the funnel and the cellophane must be watertight)
- short length of rubber or plastic tubing

- two 1 m rules
- water manometer supported in a retort stand.

Method

1 Connect one end of the tubing to the thistle funnel and the other end to one limb of the water manometer (see figure 10.12).

2 With the closed end of the funnel pointing downwards (or as shown), immerse the funnel into the water in the cylinder.

3 Lower the funnel slowly and observe the changes that take place in the manometer levels.

Observation

The water level in the limb of the manometer connected to the funnel falls while the level in the other limb rises.

Inference

Since the difference in levels of the water in the manometer measures the excess pressure of the air in the plastic tubing and funnel over atmospheric pressure and, since that pressure is the pressure of the water at the level of the funnel, then the liquid pressure on the cellophane increases as the funnel goes lower into the cylinder.

Conclusion

That the pressure due to the liquid in the cylinder increases as the depth of the point below the surface increases

In addition, we could find out whether the pressure does vary directly with the depth below the surface. To do this we measure (see figure 10.12):

1 the depth, d, of the funnel below the water surface;

2 the difference, h, between the water levels in the water manometer (or the 'head' of water in the manometer);

3 plot a graph of h/mm against d/mm.

The graph should be a straight line of unit slope. Do you know why? Can you obtain the equation connecting d and h?

This activity provides an excellent opportunity to practise M/M and O/R/R experimental skills.

ITQ6

Plan and design an experiment to show that the pressure exerted by a liquid depends on the density of the liquid.

The same arrangement can be used to show that liquid pressure depends on the density of the liquid.

Archimedes' principle ›

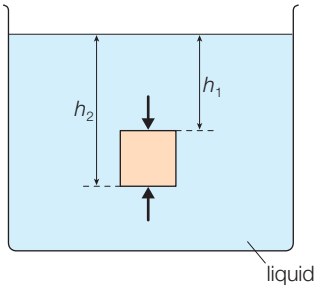
Archimedes' principle

In figure 10.13, consider a block of uniform cross-section, A, submerged in a liquid of density ρ with its top and bottom faces horizontal. There will be a force on each of these faces, since there will be a pressure of $h_1\rho g$ acting downwards on the top face and a pressure of $h_2\rho g$ acting upwards on the bottom face. So remembering that pressure = $\frac{\text{force}}{\text{area}}$, we can write:

force on top face = $h_1\rho g \times A$

and force on bottom face = $h_2\rho g \times A$

But we know that $h_2 > h_1$, so (upward force) > (downward force) and there will be a net upward force on the cylinder. This force is called an **upthrust**, but is sometimes referred to as the **buoyancy force**. It can be shown that this upthrust is equal to the weight of liquid displaced. It does not depend on the weight of the object.



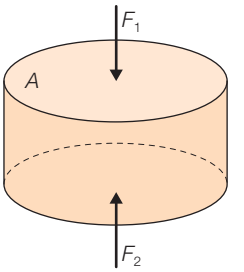


Figure 10.13

upthrust ›
buoyancy force ›

Archimedes was a Greek philosopher. He realised this principle while lying in his bath over 2000 years ago. He was so excited at detecting this upthrust that he ran out into the street shouting 'Eureka!', which means 'I have found it!' – 'it' being the answer to a problem that had been bothering him for some time.

Archimedes' principle states that a body, if immersed completely or partially in a fluid, will experience an upthrust equal to the weight of the fluid displaced.

To illustrate Archimedes' principle, we can do the following experiment.

Practical activity 10.4

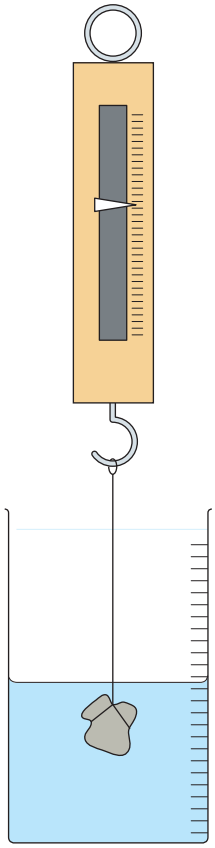


Figure 10.14

Investigating Archimedes' principle

You will need:

- newton-meter
- large measuring cylinder
- a stone that fits inside the measuring cylinder
- string.

Method

1. Tie the stone to the string and hang it from the newton-meter.
2. Note the reading on the newton-meter, N_1.
3. Partly fill the measuring cylinder with water.
4. Note the volume of water, V_1.
5. Place the measuring cylinder under the stone and raise it until the stone is completely immersed (figure 10.14).
6. Note the new reading on the newton-meter, N_2.
7. Note the new volume, V_2, of the water.

Handling the readings

What was the volume V of the stone?

$V = (V_2 - V_1)\ \text{cm}^3$

What was the upthrust (buoyancy force)?

$N = (N_1 - N_2)$ newtons

Remember: $N_2 < N_1$, because of the upthrust on the stone.

If the density of water is $1000\ \text{kg m}^{-3}$, what was the weight W of the water that the stone displaced?

mass of water displaced
$= (V_2 - V_1)\ \text{cm}^3 \times 1\ \text{g cm}^{-3}$

$= (V_2 - V_1)\ \text{g}$

so W, the weight of the water displaced,

$= \frac{(V_2 - V_1)\ \text{kg}}{1000} \times 10\ \text{N kg}^{-1}$

or $W = (V_2 - V_1) \times 10^{-2}\ \text{N}$

What do you notice about the values of N and W?

Are they equal?. They should be within the limits of experimental error, these limits depending on the precision of your readings of the newton-meter and the measuring cylinder. If the limits are such that $(N_1 - N_2)$ could lie within the limits of $(V_2 - V_1) \times 10^{-2}$, then the equality of the upthrust and the weight of the water displaced is established.

ITQ7

A newton-meter gives a reading R_1 for the weight of a metal cylinder, C, in air. A beaker containing water weighs K_1 grams on a kitchen balance. However, when the cylinder is immersed in the water contained in the beaker, its weight on the newton-meter becomes R_2 and the kitchen balance reads K_2. Explain the reason for the changes in the readings of R and K. State which of the readings R_1 and R_2 and which of the readings K_1 and K_2 is the greater.

Theoretical proof of Archimedes' principle

We can show that Archimedes' principle is true by theoretical reasoning. If we go back to figure 10.13, where the body is immersed in a liquid of density ρ, we can see that the net force acting upwards on the body $= F_2 - F_1$

$= (h_2\rho g - h_1\rho g) \times A$
$= (h_2 - h_1)\rho g A$
$= (h_2 - h_1)A\rho g$

But for a body with a uniform cross-section A, $(h_2 - h_1)A$ = volume, V, of the body.

So the net force acting upwards on the body $= V\rho g$

$= V\rho \times g$
= (mass of the liquid displaced) $\times\ g$
= weight of the liquid displaced

And so the net force acting upwards on the submerged body

= upthrust on the body
= weight of the liquid displaced by the body.

Archimedes' principle and relative density

relative density ›

The relative density of a substance is the number of times that substance is denser than water.

Archimedes' principle provides a convenient method of finding, by experiment, the relative density of a solid and also of a liquid, water being used as the reference substance. Thus, the density of aluminium is $2.7\,g\,cm^{-3}$ and that of water is $1.0\,g\,cm^{-3}$. We would therefore say that the **relative density** (R.D.) of aluminium is 2.7.

The definition of relative density is therefore:

$$\text{relative density of a substance} = \frac{\text{density of the substance}}{\text{density of water}}$$

In calculating a relative density, the individual densities must be expressed in the same unit. So relative density has no units since, in calculating its value, we divide two quantities that have identical units. The value of a relative density does not depend on the units used for the respective densities.

Worked example 10.7

A lump of a certain substance placed on a kitchen balance is shown to have a mass of 240 g. A beaker of water placed on the same balance is seen to have a mass of 420 g. When the lump of substance is suspended in the water on the balance, however, the beaker seems to weigh 480 g. Calculate the relative density of the substance.

The water in the beaker is replaced by a liquid, L, and the experiment is repeated. The mass of the lump of substance is the same as before, but the beaker with liquid now weighs 490 g and 556 g before and after the same substance is suspended in it. Calculate the relative density of the liquid, L.

Solution

When the lump is placed in the liquid there will be an upthrust on the lump. In response to this upthrust on the lump there will be a downthrust of the lump on the liquid and this downthrust is transmitted through the liquid on to the base of the container. This downthrust will be registered by the balance. Hence the apparent increase in the reading of the balance. It is as if the beaker has suffered an increase in weight. Newton's third law at work!

We have

$$\text{upthrust in water} = \frac{(480-420)}{1000} \times 10\,N = 0.60\,N$$

$$\text{upthrust in liquid L} = \frac{(556-490)}{1000} \times 10\,N = 0.66\,N$$

The upthrust is the weight of fluid displaced. We use d_W and d_L to stand for the density of water and liquid L respectively. If the volume of the substance used is $V\,cm^3$, then (since mass = density × volume)

$$\text{upthrust in water} = \frac{V \times d_W}{1000} \times 10\,N = Vd_W \times 10^{-2}\,N$$

$$\text{upthrust in liquid L} = \frac{V \times d_L}{1000} \times 10\,N = Vd_L \times 10^{-2}\,N$$

So taking the ratio we get

$$\frac{\text{upthrust in water}}{\text{upthrust in liquid L}} = \frac{Vd_W \times 10^{-2}\,N}{Vd_L \times 10^{-2}\,N} = \frac{0.60\,N}{0.66\,N}$$

or $$\frac{V\,d_W}{V\,d_L} = \frac{0.60}{0.66} \text{ (on cancelling)}$$

giving $$\text{R.D.} = \frac{d_L}{d_W} = \frac{0.60}{0.66}$$

$$= 1.1$$

The method of finding a relative density described in Worked example 10.7 is very quick and it is capable of yielding very accurate results. The instruments used for measuring masses and weights are more precise than those used for measuring volumes, and so weights and masses are more accurately measured than volumes.

Application of Archimedes' principle to sinking and floating

Sinking

An object sinks if, when completely immersed in a liquid, it moves downward in that liquid. According to Newton's Second Law, there must be a resultant or net force acting on the object. Let us see how this net force comes about. If the volume of the object is V, its density is d and that of the liquid is ρ, then

upthrust, U = weight of liquid displaced
= volume of liquid displaced × density of liquid × g
= $V \times \rho \times g$

So $U = V\rho g$

The weight of the object, W, is given by

weight, W = volume of object × density of object × g
= $V \times d \times g$
$W = Vdg$

sinking ›

For **sinking** to occur, we must have

weight of object > upthrust

which means that

$Vdg > V\rho g$

or that $d > \rho$

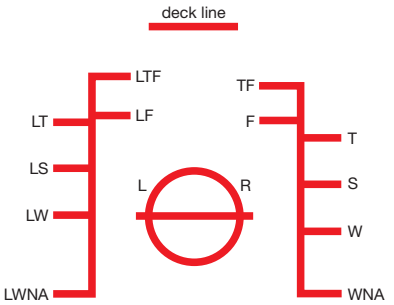
So, for an object to sink in a liquid, its density must be greater than the density of the liquid, or: density of body > density of liquid.

Thus a lump of iron (density 7.8 g cm^{-3}) will sink in water (density 1.0 g cm^{-3}) but will not sink in mercury (density 13.6 g cm^{-3}).

It should be clear, therefore, that if an object is not to sink when placed in a liquid of lesser density, it must be supported; there must be some other force acting upwards to assist the upthrust to balance the weight.

Positive and negative buoyancy

The upthrust exerted on a body which is immersed is sometimes referred to as 'buoyancy'. If the net force on the submerged body is upwards , (as in floating), it is called 'positive buoyancy'; if the net force on the body is downwards, (as in sinking) it is referred to as 'negative buoyancy'. One might say, therefore, that floating is due to positive buoyancy and sinking is due to negative buoyancy.



The letters indicate cargo, season and location:
LTF lumber, tropical, fresh
LF lumber, fresh
LT lumber, tropical
LS lumber, summer
LW lumber, winter
LWNA lumber, winter, North Atlantic
LR Lloyds Register of Shipping
TF tropical fresh watermark
F fresh watermark
T tropical load mark
S summer load mark
W winter load mark
WNA winter load line, North Atlantic

Figure 10.15 The Plimsoll line markings used on the sides of ships.

The Plimsoll line

In the nineteenth century a British politician, Samuel Plimsoll, became very concerned about the heavy loss at sea of merchant ships and trade goods. He devised a system of lines that are marked on the hull of a ship (figure 10.15), that indicate the levels at which a fully loaded ship should float. These lines indicate load levels for different waters because different kinds of water have different densities. Warm waters are less dense than colder waters, and fresher waters (such as river mouths) are less dense than salty ones. The upthrust of the water is related to its density. So a ship loaded to a safe level in cold, salty water would sink lower in warmer, fresher water.

Worked example 10.8

A small stone, of volume 20 cm^3 and density 3.0 g cm^{-3}, supported by a string, is immersed in a beaker of water (figure 10.16). Calculate the tension in the string supporting the stone.

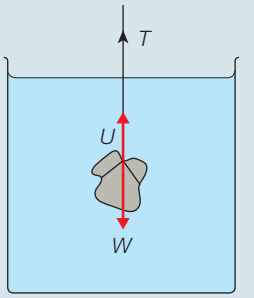


Figure 10.16

Solution

There are three forces keeping the stone at rest:

1 the weight, W, of the stone (↓);
2 the upthrust, U, of the displaced water (↑); and
3 the tension (pull), T, in the string (↑).

Since the stone is at rest

downward force = sum of upward forces

$$W = U + T$$

giving, on rearranging,

$$T = W - U$$
$$= Vdg - V\rho g$$
$$= Vg(d - \rho)$$

where $V = 20\,\text{cm}^3 = 20 \times 10^{-6}\,\text{m}^3$, $d = 3.0 \times 10^3\,\text{kg m}^{-3}$ and $\rho = 1000\,\text{kg m}^{-3}$ (all units must be S.I.). So the tension is

$$T = (20 \times 10^{-6}) \times 10 \times (3.0 - 1.0) \times 10^3\,\text{kg m}^{-3}$$
$$= 4 \times 10^{-1}\,\text{kg m}^{-3}$$
$$= 0.4\,\text{N}$$

The string supports the stone by pulling with a force of 0.4 N.

Floating

If a body does not sink in a liquid, then it either floats (remains at rest on the surface, partially submerged), or it remains suspended in the liquid without moving either upwards or downwards. If a body moves upwards to the surface of a liquid when held below the surface and then released, then there must be an unbalanced force acting upwards on it at the time of release. If we use

floating ›

a similar procedure to the one we have just used for 'sinking' to find this net force, we find that, for **floating** to occur,

density of body < density of liquid

i.e. for floating, $d < \rho$.

The principle of flotation

principle of flotation ›

The **principle of flotation** states that a floating body will displace its own weight of the liquid in which it is floating. So, for a body of total volume V and density d floating on a liquid of density ρ, with volume of the body submerged, V_{sub}:

$Vdg = V_{sub}\rho g$ (since the body is at rest)

$\frac{V_{sub}}{V} = \frac{d}{\rho}$

Since d, the density of the body, is less than ρ, the density of the liquid (condition required for floating) and $d/\rho < 1$, then $V_{sub} = d/\rho\, V$ and the volume submerged will be the fraction (d/ρ) of the volume, V, of the floating body.

It follows that if the fraction immersed is f_{imm} then

$$f_{imm} = \frac{d}{\rho} = \frac{\text{(density of the body)}}{\text{(density of the liquid)}}$$

When a body floats in a liquid, the fraction immersed,

$$f_{imm} = \frac{\text{density of the material of the body}}{\text{density of the liquid}}$$

= ratio of the densities of the body and the liquid

We may summarise, then, by saying that if the body is placed in a liquid, and:

(i) its density < the density of the liquid, the body will *float*;

(ii) its density = the density of the liquid, the body will remain at rest wherever it is placed, i.e. it will *remain suspended* in the liquid;

(iii) its density > the density of the liquid, the body will *sink*.

The principle of flotation plays an important part in the dispersal of seeds by wind action and by water action.

Boats and rafts remain afloat because of the upthrust which supports them in water. In the case of rafts the density of the logs which are strapped together to make the raft must be less than that of water. This means that the density of the wood of the logs must be less than $1\,g\,cm^{-3}$ or the relative density less than 1. This is the case for most woods found in the Caribbean. Even 'greenheart', one of Guyana's densest woods, has been floated down rivers as rafts to saw-mills where they are sawn for use in the building trade. Purpleheart, too, should be able to float as a raft, its relative density being slightly below 1.

Where submarines are concerned, they are able to sink or float as desired because they can vary their overall density by using a ballast of air or water to give them a positive buoyancy or a negative buoyancy, as required. To float, their overall relative density must be less than 1; to sink, it must be greater than 1; and to remain suspended it must be equal to 1. The submarine is designed so that when the ballast tanks are full of air the submarine just floats. If the air in the ballast tanks is slowly replaced by sea-water the submarine becomes heavier and begins to sink – the more water in the tanks, the faster the sinking. To enable the submarine to rise again, its overall density is reduced by pumping water out of the ballast tanks and replacing this water with air. This causes the overall density of the submarine to be less than that of the sea-water and the submarine rises again.

If I put a lump of metal on water, it will obviously sink since its density is greater than that of water. If I now roll the metal into a sheet and give it

the shape of a bowl, it is quite possible for that bowl to float on the water. As long as I can give the bowl such a shape that it can displace sufficient water for the displaced water to have the same weight as the bowl, the bowl will float. It seems then that the shape of the 'hull' of my bowl will be crucial to the floating or otherwise of the bowl. Is it a fact that flat -bottomed boats (like barges) will float more easily than boats with the conventional shape of a hull?

Worked example 10.9

A boiling tube of cross-section 4.5 cm^2 containing lead pellets floats in water with 10.0 cm of its length immersed. The total mass of boiling tube and pellets is 200 g. What will be the depth of immersion of the tube if a further 5.0 g of pellets is added to the boiling tube?

Solution

Let the weight of the tube at first be W_1 and the depth of immersion h_1.

Let the second weight of the tube be W_2 and the second depth of immersion be h_2.

The equilibrium equations are (i) $W_1 = Ah_1\rho g$ and (ii) $W_2 = Ah_2\rho g$

Subtracting the equations, we have $W_2 - W_1 = (h_2 - h_1)A\rho g$

and so $h_2 - h_1 = \dfrac{(W_2 - W_1)}{A\rho g}$

We recognise that the change in immersion depth = $h_2 - h_1$ and that the extra weight = $W_2 - W_1$.

So remembering that all units must be S.I. and substituting, we have

$$h_2 - h_1 = \frac{5.0 \times 10^{-3} \times 10}{4.5 \times 10^{-4} \times 10^3 \times 10}$$
$$= 0.0111 \text{ m}$$
$$= 1.1 \text{ cm}$$

Therefore the additional depth of immersion = 1.1 cm and the new depth of immersion will therefore be 11.1 cm.

Chapter summary

- Pressure on a surface is the force acting at right angles to that surface over unit area. This area is the area of contact between the surface and the agency responsible for the force.
- The general formula for pressure is
  $$\text{pressure} = \frac{\text{force acting perpendicular to the surface}}{\text{area of contact between agency and surface}}$$
- The S.I. unit of pressure is the newton per metre squared or the pascal (Pa).
- A solid exerts a pressure on a surface when either the weight of the solid causes a thrust to be exerted on that surface, or the solid transfers a thrust to the surface.
- The pressure of a liquid is due to the perpendicular thrust exerted on a surface placed in the liquid.
- The formula for the pressure due to a stationary liquid on a surface or at a point in that liquid is
  pressure = (height of liquid above surface or point) × (density of the liquid) × (gravitational field strength)
  or $P = h\rho g$

- The pressure of a fluid acts at all points in that fluid and in all directions about that point and towards the point. A fluid is any liquid or gas.
- The hydrostatic pressure at all points on a given horizontal level in a stationary, continuous liquid is constant.
- Pressure applied at any point in an enclosed liquid or gas is transmitted undiminished in value to all points within that liquid or gas. This is known as Pascal's principle.
- The fact that the pressure of a liquid or a gas varies with depth explains why a body that is immersed partially or wholly in a fluid experiences an upward force, called an upthrust.
- Archimedes' principle states that a body partially or completely immersed in a fluid experiences an upthrust that is equal to the weight of the fluid displaced by the body.
- Archimedes' principle leads to the principle of flotation. This principle states that a floating body displaces its own weight of the liquid in which it is floating.
- Archimedes' principle provides a very convenient method of finding experimentally the relative density of a solid body or of a liquid.

Answers to ITQs

ITQ1 **(a)** 2000 N, **(b)** 3000 Pa

ITQ2 250 N

ITQ3 17.7 kPa

ITQ4 **(i)** Alter the angle between the glass tube and the thistle funnel between 0° and 90° by heating the glass. Use at least five different angles.
(ii) For each angle between glass tube and the funnel and for a given depth below the surface, measure the pressure with a water or oil manometer.
(iii) These values of the pressure should all be the same within the limits of error, the errors being those involved in measuring the depth of the funnel (this could be a large error) and the head of liquid in the manometer.

ITQ5 **(i)** The R.D of mercury is large. So for the same pressure to be measured, the head will be shorter for mercury than for all other common liquids, since they all have a much smaller density than mercury.

ITQ6 **(i)** Select three or four liquids which can be obtained in large quantities (like water, kerosene, motor oil and one other), the densities of which are known.
Then either
A **(ii)** measure the pressure at widely different depths in each of the liquids using a water or an oil manometer;
(iii) plot a graph of pressure against depth of the liquid for each liquid used. Points on each graph should give a straight line, since for each set of measurements made the density of the liquid was constant. The slope of the graph would be the density of the liquid.
(iv) Different slopes would suggest that the density is important and that, for a given depth below the surface, the pressure of the liquid depends on the density of the liquid, but you would not know how.
Or
B (Which is the better method) Do you know why?
(ii) measure the pressure at given depths in the different liquids;
(iii) plot a graph of pressure against density.
(iv) If the graph is straight, the conclusion would be that the pressure in the liquid is proportional to the density if the depth in the liquid is kept constant.

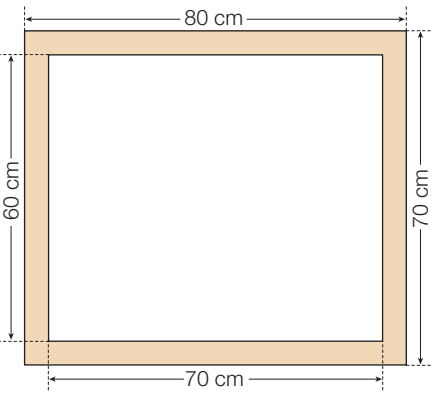
ITQ7 $R_2 < R_1$ because of the upthrust on the cylinder and therefore on the newton-meter; $K_2 > K_1$ because of the downthrust in response to the initial upthrust on the cylinder. The downthrust is transmitted to the base of the container and, of course, down to the balance pan.

Examination-style questions

1 (i) The torr is the unit of pressure most often used for expressing the pressure of a partial vacuum. It may be defined as the pressure exerted by a column of mercury 1 mm high. Calculate an approximate value for the torr in pascals. Look up the origin of the unit 'torr'. It is a contraction of the name of a past scientist. Who was he? What was his significant scientific contribution?

(ii) The bar is defined as 10^5 Pa. Calculate the ratio: 1 millibar/1 torr.

2 A writing desk stands on a base of external dimensions 80 cm × 70 cm and internal dimensions 70 cm × 50 cm (see diagram). The total weight of the desk and base is 400 N. Calculate the average pressure the desk exerts on a horizontal floor, (i) in N cm^{-2}, and (ii) in Pa.



3 A heavy lorry stands on six tyres, four at the back and two in front. The total weight of the lorry is 40 000 N. The area of contact between each rear tyre and the ground is 100 cm^2 and that between each front tyre and the ground is 80 cm^2. Assuming that the pressure is the same in all the tyres, calculate the value of this pressure.

4 Explain each of the following:

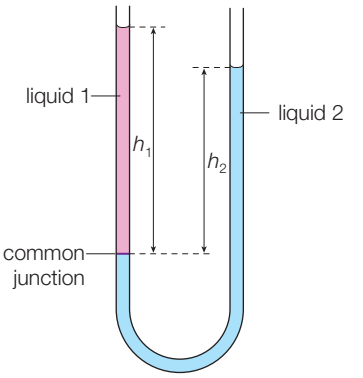
(i) If a drinking glass is filled to the brim with water and a flat sheet of stiff plastic is slid over the top and the glass inverted, the plastic sheet remains in place only if there is no air bubble in the water.

(ii) Heavy vehicles can move on soft ground without becoming stuck only if their tyres are very wide.

(iii) A sharp knife will cut hard objects more easily than a dull one.

(iv) A heavy haversack is more comfortable to carry over the shoulder if the straps are wide than if they are narrow.

5 The following is one method of comparing the densities of two immiscible liquids directly (see diagram). Stand a tall U-tube upright in a retort stand. Use a small funnel to pour the denser liquid into one limb of the U-tube to a height of about 15 cm. Now slightly incline the tube and, again using the funnel (having first washed and rinsed it), pour the other liquid down the other limb. Return the U-tube to its vertical position. When the liquids have settled, measure the vertical height of each liquid from the common junction. Denote these heights h_1 and h_2 as shown in the diagram.



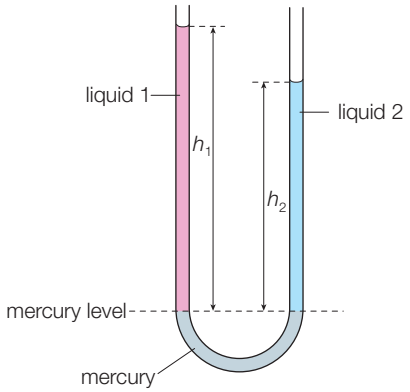
(i) Using Pascal's 'law' (pressures on the same horizontal level in a stationary liquid are the same), obtain an equation connecting the heights h_1 and h_2 and the corresponding densities ρ_1 and ρ_2.

(ii) Use this relation to find the density of an unknown oil, where h_1 (for oil) = 33.4 cm and h_2 (for water) = 28.6 cm.

This method is called the 'balancing columns' method. It can be used to find the ratio of the densities of liquid 1 (like an oil) and liquid 2 (water). This ratio is then the relative density of liquid 1. This method can be used only if the liquids are immiscible.

6 If the liquids in the previous question are miscible, a short length of mercury is introduced into the U-tube to separate them, as shown in the diagram.

The mercury menisci are both on the same horizontal. Use Pascal's law to obtain an equation connecting the heights of the liquids and their densities.



7 In another method, called Hare's method, the U-tube is inverted and the ends are dipped into two liquids to be compared. A length of rubber tubing fitted to an opening at the bend in the tube is used to suck the two liquids up the tubes.

(i) Explain why the pressures due to these liquids at the levels of the menisci in the beakers (levels P and Q) are equal. Note that these levels need not be the same.

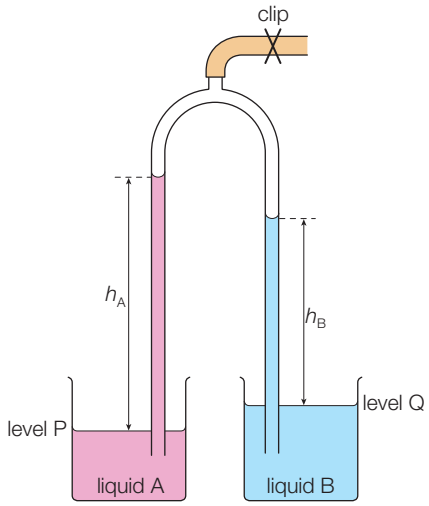
(ii) To what agency are the pressures over the menisci in the two beakers due?

(iii) To what agency are the pressures at the same levels (i.e. the level of the meniscus in the beakers), but with the tubes, due?

(iv) Use Pascal's law to write expressions for these pressures.

(v) Hence obtain an equation connecting the pressures.

(vi) Now deduce an expression for the ratio of the densities of the liquids A and B.



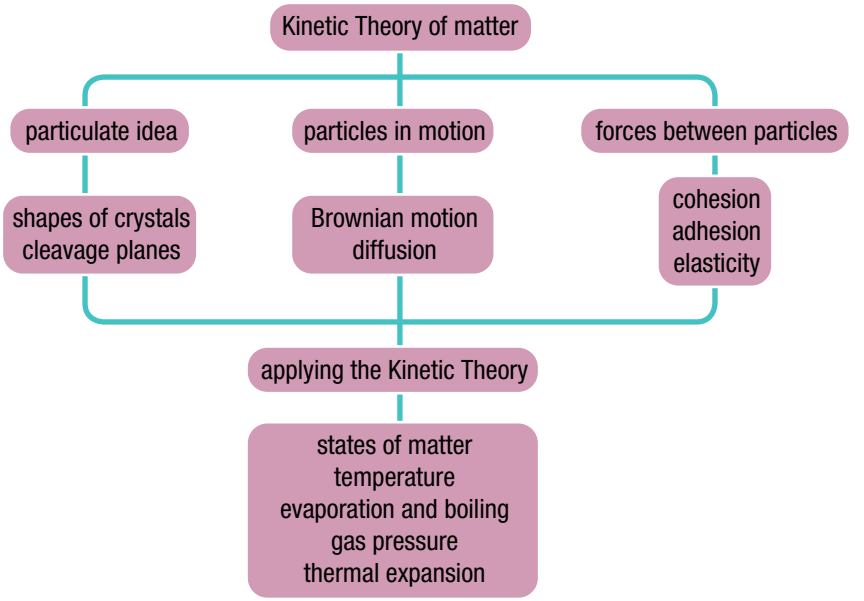
Section B:

Kinetic Theory and Thermal Physics

11 The Kinetic Model of Matter

By the end of this chapter you should be able to:

- state evidence that matter is made up of particles
- cite evidence that the particles of matter are moving (Brownian motion)
- cite evidence for the existence of forces between particles
- use the Kinetic Theory to explain differences in the macroscopic properties of solids, liquids and gases (vapours)
- use the Kinetic Theory to explain simple common phenomena, such as evaporation and boiling
- demonstrate that temperature remains constant during a phase change;
- give examples of the cooling effect of evaporation
- explain gas pressure in terms of molecular motion
- relate the temperature of a body to the average kinetic energy of its molecules
- use the Kinetic Theory to account for thermal expansion of solids, liquids and gases
- explain everyday observations of the effects of thermal expansion
- discuss useful applications of thermal expansion



Particles of matter

An element consists of one kind of matter only (e.g. hydrogen, copper, mercury).

atom ›

An **atom** is the smallest particle of an element that retains the characteristics of the element.

molecule ›

The smallest particle of a compound or an element that can exist by itself is called a **molecule**.

ion ›

An **ion** is an atom or molecule that has gained or lost an electron.

The Kinetic Theory of matter

Kinetic Theory

The kinetic model of matter is based on the **Kinetic Theory**, which makes the following assumptions about matter:

- Matter is made up of tiny particles, such as molecules (figure 11.1).
- The particles of matter are always in motion.
- There are strong forces between particles separated by very short distances.
- Collisions between these particles are perfectly **elastic**. This means that, overall, energy is not lost by particles during collisions.

elastic

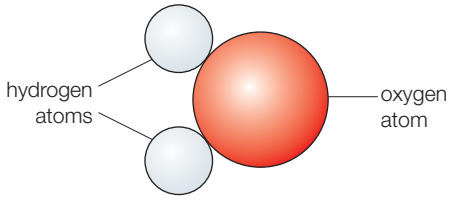


Figure 11.1 A model of a water molecule. A molecule of water is made up of smaller particles called atoms. Some atoms (e.g. the inert gases) can exist by themselves in Nature. Others cannot; they are combined with other atoms. We will look more closely at atoms in section F.

Only certain types of substances consist of molecules. A solid conductor does not consist of molecules since it has a structure of a vibrating lattice of ions through which electrons meander freely. An ionic solid (e.g. sodium chloride) consists of a lattice arrangement of positive (Na^+) and negative (Cl^-) ions. The kinetic model of matter applies to these types of substances as well.

In this chapter, we are going to examine the evidence for the kinetic model of matter. We will then use the model to account for elastic behaviour, states of matter, evaporation, boiling, gas pressure, temperature, and expansion of matter when heated.

Evidence for the Kinetic Theory

The word 'kinetic' stems from the Greek word *kinema*, which means 'motion'. The Kinetic Theory is therefore a theory involving motion. The theory states that matter is made up of particles which are always in motion and that there are forces between these particles. There are several pieces of evidence that support the Kinetic Theory of matter. The evidence, though, is indirect.

Evidence that matter is made of particles

Shapes of solids

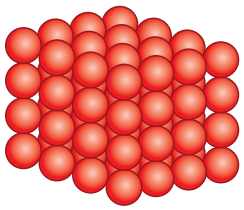


Figure 11.2 A cuboid shape formed from an orderly arrangement of spheres.

Figure 11.2 shows a cuboid shape formed from rectangular arrangements of spheres. If we imagine the spheres to be very small 'particles', the picture shows how flat, smooth surfaces can be formed from orderly arrangements of particles.

Practical activity 11.1

Examining grains of substances

Using a magnifying lens or a microscope, observe and draw one *grain* of each of the following substances: salt (sodium chloride), sugar, potassium permanganate (potassium manganate (VII)) and corn meal.

Questions

1 Which of the grains had smooth, flat surfaces? Which did not?

2 What do you infer from these observations?

Cleavage planes

cleavage planes

Figure 11.3 shows a calcite crystal with the edge of a sharp blade placed on its surface. If the blade is tapped lightly when parallel to an edge of the crystal, the latter splits into two. Smooth surfaces, called **cleavage planes**, are formed at the split. If the blade is tapped when its edge is not parallel to

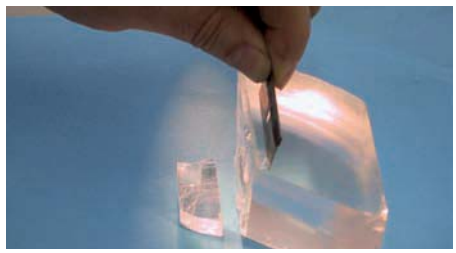


Figure 11.3 The splitting of a calcite crystal, showing a cleavage plane.

the edge of the crystal, various sizes of irregularly shaped crystal fragments are produced.

The formation of smooth cleavage planes only when the blade is parallel to an edge suggests a regular arrangement of particles within the crystal. The space between two rows of particles represents a weak area. The crystal is easily split into two planes when a force is applied (e.g. using a sharp blade) along the space.

If the particles of a substance are arranged in a regular manner, the substance is said to have a **crystalline** structure.

crystalline ›

Practical activity 11.2



Investigating ice for cleavage planes

You are advised to wear safety goggles for this activity.

Your teacher will most likely carry out this activity. If he or she does, you are to observe the following demonstration and answer the questions below (it is not advisable for the student to carry out this activity). The teacher will:

1. Place the sharp edges of a knife on the top surface of an ice cube.
2. Using the hammer tap *lightly* on the blunt edge of the knife, keeping the blade of the knife perpendicular to the surface of the cube.
3. Repeat the light tapping of step 2 using pieces of the ice formed. (The knife blade will be oriented parallel to, perpendicular to, as well as various other angles to, the position used in step 2.)

Questions

1. Describe the cleavage planes formed when the ice broke exactly into two pieces.
2. If flat cleavage planes were formed, how does this support the idea that ice is made up of particles?
3. If cleavage did not result in perfectly flat planes, does this contradict the particle theory? Explain your answer.

Extension

Snowflakes show regular shapes with straight edges. Find out and sketch what some of these shapes look like. Can you suggest how these shapes are further evidence that matter is made up of particles?

Evidence that particles move

Diffusion

The scent of a perfume spreads in all directions through the air. If perfume is made up of particles, then this behaviour suggets that the particles must be moving. Also, a purple colour spreads when a crystal of potassium permanganate (potassium manganate (VII)) is placed in water (figure 11.4). This suggests that coloured pieces (particles) are breaking off the solid crystal and are making their way through the water. Such movement of particles is called **diffusion**.

diffusion ›

Diffusion is the random movement of particles in all directions. In diffusion, more particles move from a region of high concentration to one of lower concentration than the other way round.

Diffusion takes place randomly in all directions. That is why, after a few days, the colour in the beaker in figure 11.4 becomes uniform. More particles move from a region of high concentration to one of lower concentration than the other way round. The overall effect of the random motion is, therefore, that a uniform mixture is formed.

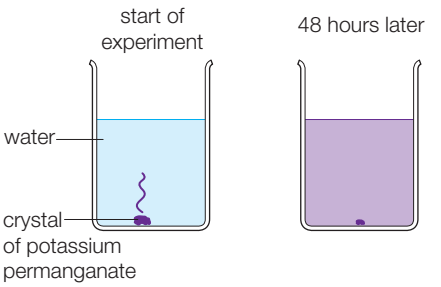


Figure 11.4 Diffusion of potassium manganate (VII) in water.

BIOLOGY & CHEMISTRY: diffusion & osmosis

Practical activity 11.3

Observing diffusion in a liquid

You will need:

- a few crystals of potassium permanganate (potassium manganate (VII)), preferably at least 1 mm long
- 250 ml measuring cylinder.

Method

1 Pour water in the measuring cylinder to the 250 ml mark.

2 Drop the crystals into the cylinder of water. Leave the set-up undisturbed.

3 Observe the spreading of the colour in the cylinder over a period of 3 days at intervals of a few hours. At each observation, note the time and sketch the appearance of the colour pattern.

4 Note how long it takes for the crystal to 'disappear'.

Questions

1 What do the gradual disappearance of the crystal and the consequent appearance of the colour in the liquid suggest about the make-up of the crystal?

2 What does the spreading of the colour suggest about the particles of potassium permanganate in the liquid?

Extension

Would the colour from crystals of copper sulphate spread at the same rate in water as that from potassium permanganate? How would you design an experiment to investigate this?

ITQ1

Perfume diffuses through a gas (air). Potassium manganate (VII) diffuses through a liquid (water). Can substances diffuse through a solid?

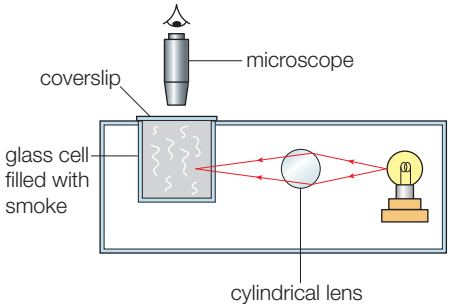


Figure 11.5 Use of a smoke-cell to observe Brownian motion.

Brownian motion ›

Brownian motion

When observing pollen grains under a microscope, a 19th-century Scottish biologist, Robert Brown, noticed that the pollen grains, suspended in water, were trembling and moving in a random jerky way in various directions. Brown reasoned that particles (molecules) of water were probably in motion and colliding with the pollen grains. The random movement of the pollen grains became known as **Brownian motion**.

Figure 11.5 shows a smoke-cell (a transparent box containing smoke) in which Brownian motion is observed in a gas (air). Bright specks showing jerky movement are observed through the microscope. The bright specks are small grains of soot that are illuminated by the strong light.

ITQ2

How is the jerky movement of the specks of soot evidence that air is made up of tiny particles (molecules)?

Evidence that there are forces between particles

Clean, dry microscope slides tend to stick to one another. If the slides are made up of particles then this suggests that the particles are attracting each other. (When particles of the same kind attract one another, we call the force between them '**cohesion**'.) The fact that the slides do not stick together if tiny bits of dirt lie between them suggests that cohesive forces act over extremely tiny distances only.

cohesion ›

Water also tends to stick to a clean microscope slide. This suggests that the particles of glass and water are attracting one another. (The force of attraction between dissimilar particles, e.g. between glass and water, is called '**adhesion**'.)

adhesion ›

ITQ3

A drop of water on a clean glass surface will spread out, but a drop on a leaf will stay roughly spherical (figure 11.6). How can one explain this in terms of forces between particles?

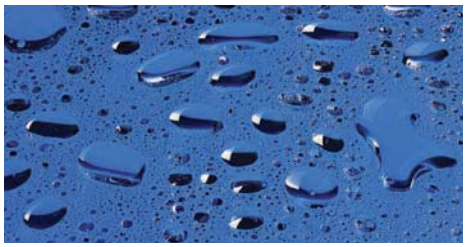




Figure 11.6 Appearance of drops of water placed on a glass surface and on the top surface of a leaf.

elastic behaviour ›

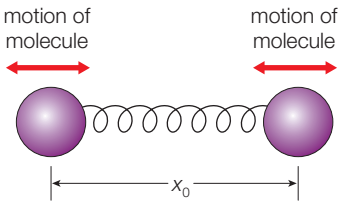


Figure 11.7 For separation distances greater than x_0 the force between the molecules is attractive, as though a 'spring' is pulling them together. For separation distances less than x_0 the force between the molecules is repulsive, as though a compressed 'spring' were pushing them apart. However, the force of attraction weakens rapidly to practically zero when the separation is much larger than x_0. The spring model applies only to molecules that are close together.

temperature ›

pressure ›

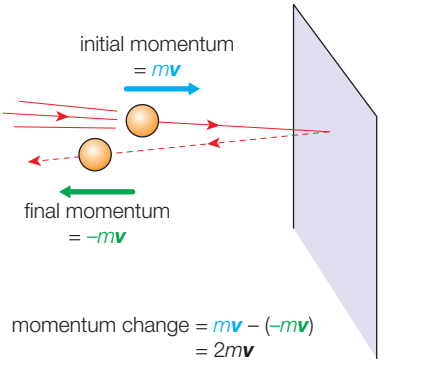


Figure 11.8 Magnitude of momentum change of a particle as it bounces off a wall.

CHEMISTRY: particulate nature of matter

molecular theory ›

Wires, strings and rubber bands resist being stretched when pulled from both ends. When the applied pulling forces are removed, these solids return to their original shapes and sizes (providing the forces have not been too large). This **elastic behaviour** suggests that there are strong forces of attraction between particles that make up the materials.

Solids also resist being compressed. So do liquids. The latter can be shown using liquid in a syringe, by blocking the exit of the liquid and applying a force to the plunger of the syringe. These two examples suggest that particles in solids and liquids repel each other when an attempt is made to bring them closer together.

The particles of solids and liquids therefore behave as though there were invisible 'springs' between them (figure 11.7).

Applying the Kinetic Theory

The Kinetic Theory can be used to explain some of the macroscopic properties and behaviours of matter, such as temperature, pressure, states of matter, change of state and thermal expansion. The explanations are discussed in the sections that follow.

Temperature

The degree of 'hotness' or 'coldness' of an object, which are both relative terms, is called its **temperature**. The Kinetic Theory, however, does not associate temperature with the highly subjective impression of hotness or of coldness. It relates temperature to the average kinetic energy of the particles of an object. Thus, at high temperatures, particles are moving at high speeds. At low temperatures, they are moving at low speeds. At the absolute zero of temperature, particle motion would cease.

Gas pressure and volume

The Kinetic Theory can also account for the **pressure** that a gas, or a vapour, exerts on the walls of its container. The pressure exists because a gas consists of molecules that are traveling randomly at high speeds and thus are constantly colliding with the walls of their containers. To put this in more precise terms, there is a change in direction of the velocity of a particle when it bounces off a surface, so there is a momentum change (see figure 11.8). As described by Newton's Second Law (see page 113) the steady changing of momentum by the particles results in a force on the surface, observed as a pressure on the walls of the container.

Suppose we were to reduce the volume of the container without letting any gas escape. The particles would now travel shorter distances before colliding with the walls. There would be more collisions on the wall each second and, hence, a great rate of change of momentum, a greater force on the container walls and, as a result, a greater gas pressure. We would need, therefore, to apply an increased force on the gas in order to reduce the volume it occupies. The topic of gas pressure will be treated further in chapter 15.

States of matter: solids, liquids and gases (vapours)

Solids

The **molecular theory** suggests that, in a solid, the particles are close together (figure 11.9) and that, at such close distances, the force of attraction between

Metallic solids (e.g. wires) are not made up of individual molecules. Hence a model, different from the molecular model, is used to account for the forces of attraction and repulsion between particles.

particles is large. The fixed shape of solids can be explained on the basis of this powerful force of attraction: the particles cannot move away from one another because the attractive forces between them are so large.

If an attempt is made to pull apart a solid (e.g. a thread), the force needed becomes greater the more the thread stretches. This observation suggests that:

- there are attractive forces acting between the molecules of the thread which try to prevent the thread from stretching; and
- these forces increase as an attempt is made to get the molecules further apart (figure 11.7).

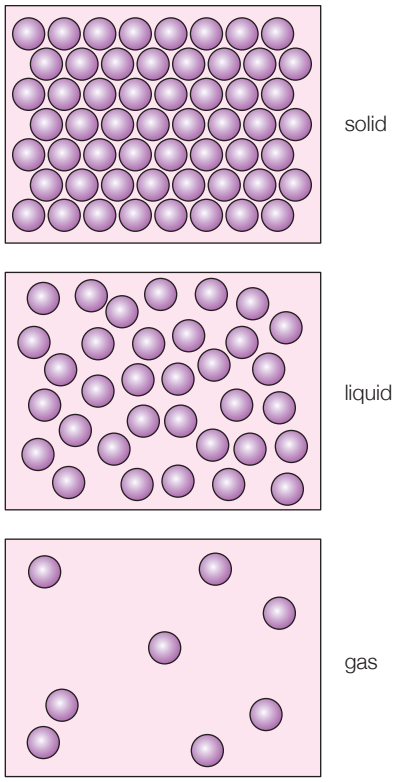


Figure 11.9 Particle model of solid, liquid and gas.

incompressible ›

If a solid (e.g. a wooden block) is squeezed, the volume hardly changes at all. We say that solids are **incompressible**. The volume changes so little because the particles move scarcely any closer to one another. This suggests that there are very large repulsive forces between particles as well as the attractive forces (figures 11.7 and 11.9).

Characteristics of a solid: fixed shape; fixed volume; incompressible.

The fixed shape of solids suggests that the attractive and repulsive forces between neighbouring molecules balance each other at a set distance apart. Since, according to the Kinetic Theory, the particles are always in motion, each particle **vibrates** about a fixed position (figures 11.7 and 11.9).

vibrates ›

Liquids

amplitude ›

When a solid is given sufficient heat energy, the solid may change to a liquid. The particles vibrate with greater speed, and with larger **amplitude**. The mean separation distance between the particles increases, though not by much, as shown in figure 11.9. The separation distances are large enough, however, to cause the force of attraction between particles to be *reduced*. (Remember that, in the 'spring model' of forces between molecules, the force of attraction at first increases with distance between molecules, but then weakens when the separation increase is too large. The spring model applies only to molecules that are close together.)

Generally, because the force of attraction between the molecules is now so weak (as compared with the force of attraction between particles of a solid), the particles in a liquid are easily moved relative to one another. The shape of a substance in the liquid state may therefore be easily changed. Thus, a liquid will 'pour' easily, taking the shape of the container into which it is poured (figure 11.10). The liquid will, however, have a definite volume since the force is still strong enough to keep

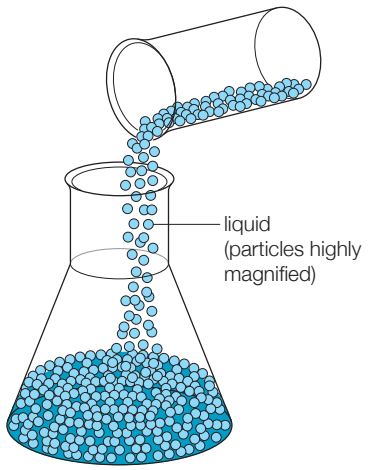


Figure 11.10 A liquid pours easily because the force of attraction between particles is not as strong as that in solids. A liquid also takes the shape of the container into which it is poured.

the particles together, causing the space the liquid occupies (or its volume) to remain the same.

Any attempt to squeeze a liquid results in very little change in volume of the liquid. This is because the particles are still fairly close to each other and the repulsive force between them is still large. Thus, like solids, liquids are incompressible.

Characteristics of a liquid: pours easily, taking the shape of the container into which it is poured; fixed volume; incompressible.

Gases (vapours)

When liquids are heated, the particles may receive enough kinetic energy (motion energy) to move away from one another completely. The average separation distance between particles is now so large that there is hardly any force of attraction between them. The particles now move *independently* of one another at high speed, and in many different directions. In the gaseous state, the particles are moving (and existing) independently. Hence, they are molecules.

Particles can travel throughout the entire volume of any container until they hit the walls of the container. (The particles also collide with one another.) Thus a gas will fill any size or shape of container into which it is put. Therefore, a gas, like a liquid, has no fixed shape, but unlike a liquid, will not retain a particular volume.

Characteristics of a gas: fills any size or shape of container into which it is put; has no fixed shape, but unlike a liquid, has no fixed volume; compressible.

Since the mean separation distance between molecules is very large, the force of repulsion between molecules is also small. Hence, molecules of a gas can be brought closer to each other easily by application of a force, for example in a bicycle pump. An increase in the force on the piston results in a decrease in volume of the gas. This shows that gases are *compressible*.

Practical activity 11.4

Compressibility (qualitative observation)

Compare the changes in volume when:

1. a force is exerted by a plunger on air trapped in a tightly sealed gas syringe;
2. a force is exerted by a plunger on water trapped in a tightly sealed syringe;
3. a compressive force is exerted between the ends of a piece of wood.

Change of state

During a change of state of a pure substance, for example from solid to liquid or from liquid to vapour, the temperature of the substance remains constant.

melting point ›

The temperature at which a solid becomes liquid at standard pressure is called the **melting point** of the solid. When the liquid is heated further, boiling takes place – bubbles of vapour are formed within the liquid. Throughout boiling, the temperature remains constant. The temperature at which boiling takes place at standard pressure is called the **boiling point** of the liquid. After all the liquid has been converted to vapour, the temperature of the vapour rises as heat is added to it.

boiling point ›

When the liquid is cooled, the temperature again becomes steady during the change from vapour back to liquid. The change from vapour back to liquid is called condensation. As the liquid is further cooled, the temperature again becomes steady during the change from liquid back to solid. The change from liquid to solid is called freezing.

Figure 11.11 shows the heating of candle wax in a water bath. Since the vapour of candle wax is flammable, it is not recommended that this activity be done in school.

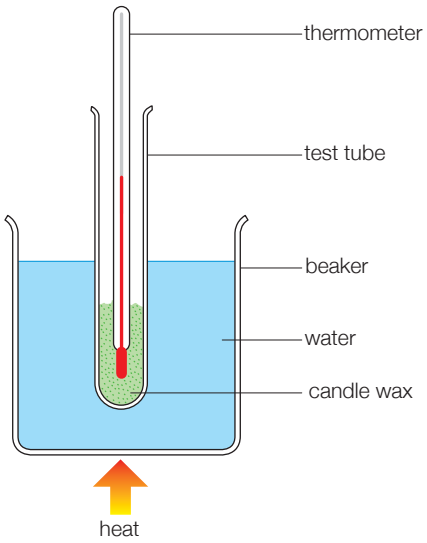


Figure 11.11 Heating crushed candle wax in a water bath.

change of state ›

ITQ4

(a) How do you explain the horizontal shape of the graph (figure 11.12) as solid candle wax changes to liquid?

(b) What if there was no distinct horizontal shape in the graph during the change from solid to liquid – how would you explain that?

Latent heat

When heat is applied to a pure substance undergoing a change of state, a temperature change of the substance does not manifest itself. The heat involved is called 'latent (hidden) heat' since its application is not manifested by a change in temperature. The topics of latent heats and changes of state are taken up in more detail in chapter 13.

Volatile liquids

A liquid is said to be volatile if it evaporates readily at or below room temperature (e.g. chloroform, Freon, methanol).

A graph of temperature against time shows that the temperature of the solid rises and then becomes steady (figure 11.12 (a)). It is at this steady temperature that a change of state occurs (from solid to liquid, in this case). When all the solid has been converted into liquid, the temperature of the wax, now liquid, starts to rise again until the temperature of the water bath is reached.

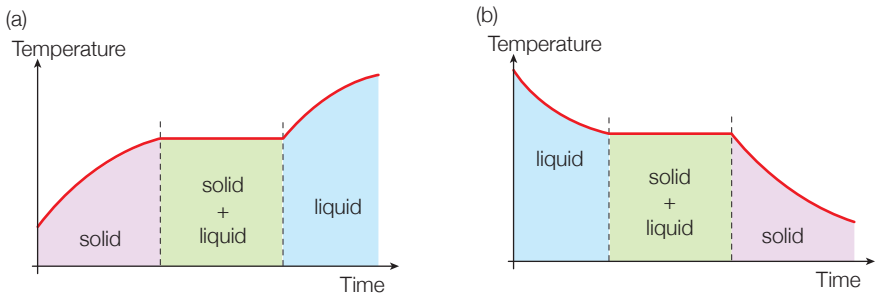


Figure 11.12 Solid–liquid cooling curves: graphs of temperature against time for candle wax during: (a) heating, and (b) cooling.

During a **change of state**, the average separation distance of the particles within a substance changes. In the case of the wax, as the solid is heated, the average distance between the wax molecules increases. The heat energy supplied causes molecules of solid wax to increase their speed and amplitude of vibration. This results in an increase in the overall energy of the molecules, which is both kinetic and potential (stored) energy. However, the energy supplied during the change of state does not increase the kinetic energy of the molecules. Rather, it only does work in *increasing* the separation of the molecules. That is why the temperature is *constant* during the change from solid to liquid (i.e. during melting).

When the liquid wax is cooled, the temperature again becomes steady during the change from liquid back to solid (called 'freezing'). The freezing temperature, or freezing point, is the same as the melting point of the wax, at a given external pressure. During the change from liquid to solid, the average separation distance between molecules *decreases* and energy (latent heat) is given up to the surroundings, while the temperature still remains constant.

At a certain temperature, which is higher than the boiling point of water, liquid wax begins to change into vapour by boiling. (**Caution**! Don't try it!) Throughout boiling, the temperature remains *constant*. At this temperature, the heat energy being supplied does work in effecting a very large increase in average separation distance between molecules as liquid wax changes to vapour. (The reverse of the process of boiling occurs during condensation, when the vapour becomes liquid upon cooling, and gives up heat without its temperature changing.) When all the liquid has changed to vapour, the temperature of the vapour rises as heat energy continues to be supplied.

Evaporation

The change from liquid to vapour that takes place at the surface of a liquid is called evaporation.

The Kinetic Theory explains evaporation as follows. In a liquid, the particles are moving about with a variety of speeds, and therefore with a variety of kinetic energies, even though the liquid as a whole is at a uniform temperature. Some of the faster particles, when present at the surface, have enough energy to enable them to break loose from the liquid and move independently above the liquid as a vapour. Thus evaporation can take place *at any temperature*. The rate at which evaporation takes place increases as

ITQ5

When faster particles escape from the surface of a liquid, as occurs during evaporation, they take kinetic energy with them. What happens to the average kinetic energy of the particles remaining in the liquid when evaporation occurs? What happens to the temperature of the liquid?

temperature increases, since, at high temperatures, more of the particles have the energy needed to break free from the liquid.

Cooling effect of evaporation

The answer to ITQ5 indicates that evaporation is accompanied by a *decrease* in the average kinetic energy of the particles of a liquid. This results in a drop in the *temperature* of the liquid since temperature decreases with decreasing average kinetic energy of particles. Thus, evaporation results in a cooling effect in a liquid.

Group discussion and activity 11.5

Observing the cooling effect of evaporation



Caution: Alcohol is highly flammable – do not use near a naked flame or a hot object.



You will have access to the following equipment: rubbing alcohol, thermometers, strips of cloth, retort stands with clamps, small beaker with water.

You are to design and carry out activities to investigate:

1. the cooling effect produced by the evaporation of rubbing alcohol on the skin (both with and without blowing on the alcohol when it is on the skin); and
2. the cooling effect produced by evaporation of water on a piece of cloth.

You are to present your plan to your teacher *before* carrying out the activities. Your plan must include controlling of variables.

Extension

The cooling effect produced when a liquid evaporates is applied in the evaporator in a refrigerator and in the hygrometer used in weather forecasting. Find out how each of these items works.

Also, find out the brand names of liquids that are advertised as keeping the body cool, especially during a fever. How is the cooling effect produced when these liquids are used?



Figure 11.13 Water in a plastic pitcher and in an earthenware jar.

ITQ6

Water at room temperature is placed in a room in a closed non-porous, plastic pitcher (see figure 11.13). Would the temperature of water drop to a few degrees below room temperature? Explain your answer.

Old-fashioned (unglazed) earthenware jars can actually keep water at a temperature below room temperature through evaporation from their surfaces. The jars are porous, and so water seeps slowly outwards through the fine pores in the walls of the vessels. The water, on reaching the outer surface of the jar, evaporates steadily, especially if there is a strong draught blowing past the jars. This is why in years gone by, especially in Guyana, these jars (called 'goblets') were placed in 'coolers' (shaded areas near open windows). During the evaporation, heat was taken from the jar and the water inside it. Thus the temperature of water dropped to a few degrees below room temperature. To find out what happens with plastic pitchers, see ITQ6.

The cooling effect of sweating is also explained in terms of evaporation. When sweat evaporates, heat energy is taken from our skin, which is in contact with the sweat, so we feel cooler.

BIOLOGY: homeostasis

Factors affecting evaporation

If evaporation takes place in a closed container, molecules of vapour in their random movement will return to the liquid as well as escape from the liquid. A **dynamic equilibrium** is set up when as many molecules of the liquid are

dynamic equilibrium ›

escaping as those of the vapour are returning. In an open container, more molecules are likely to leave the liquid than to return to it and so the volume of liquid slowly reduces. In the case of a covered vessel this does not happen.

Practical activity 11.6

Rate of evaporation

You are to design and carry out *one* investigation in *each* case to determine how the rate of evaporation is affected by:

1 surface area;
2 presence of wind or draught;
3 temperature.

As usual, you must pay strict attention to control of variables, and you must show your plans to your teacher before attempting to carry them out.

ITQ7

Explain, on the basis of the molecular theory, why wet clothes hung on a line dry faster on dry days than on humid days at the same temperature.

Boiling

vapour pressure

When a liquid boils, some molecules gain enough energy to form a vapour within the liquid. Bubbles containing this vapour are seen rising. If the pressure of the vapour within the bubbles, called the **vapour pressure**, is less than the pressure outside, the bubbles collapse. (The outside pressure is usually due to the atmosphere, and will vary.) So if the bubbles are to survive and rise to the top of the liquid, the vapour pressure inside them must be at least equal to the outside pressure.

boiling

For boiling to occur, the vapour pressure inside the bubbles must be equal to the pressure outside.

Note that, whereas evaporation may take place at any temperature, boiling will take place at a specific temperature for a particular value of external pressure.

To maintain boiling, energy must be supplied continuously to the liquid. Once boiling starts, the temperature of the liquid does not change even though heat energy is being supplied. The energy supplied is used to do work in moving molecules of the liquid far apart from each other, in changing the state from liquid to vapour.

Factors affecting the boiling point of a liquid

The external pressure

The temperature at which boiling takes place, called the boiling point, depends on external pressure. We have noted that boiling occurs when the pressure of the vapour formed inside the bubbles is equal to the external pressure. Suppose that there is an increase in the external pressure. For the liquid to boil, the vapour pressure inside the bubbles will now have to be higher, and this will only be possible if the molecules are moving faster; in other words, if the vapour is hotter. So boiling will now happen at a higher temperature.

A pressure cooker (figure 11.14) enables food to be cooked more quickly because the water in the cooker can be heated to much higher temperatures than usual. Inside the pressure cooker, the water vapour produced by evaporation during heating creates a very high pressure in the space above the water, since the vapour is not allowed to escape. The pressure inside the cooker rises so much that the vapour pressure of any bubbles that form does not equal that of the vapour in the space. Bubbles do not form in the cooker, and there is no boiling. Thus the temperature of the water, and food, in the cooker continues to rise much higher than the temperatures found in an ordinary pan, and so the food is cooked more rapidly. The pressure inside the cooker, and so, also, the temperature, is controlled by a 'weight' which rises at a certain pressure, opening a valve, which allows very hot vapour to leave the cooker. In this way, the temperature reached by the cooker can be controlled.

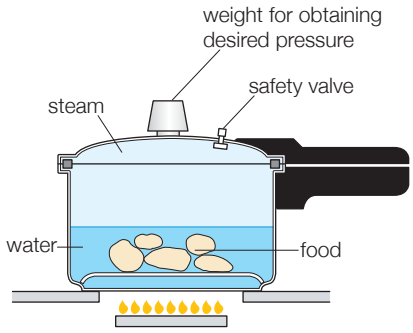


Figure 11.14 A pressure cooker. Note the safety valve, which releases steam when the pressure inside the cooker gets dangerously high.

On Mount Everest (which is approximately 9000 m high) water boils at 70°C since the air pressure there is about one-third of that at sea level.

ITQ8

Why does water boil at temperatures less than 100°C when the atmospheric pressure is less than 1 atmosphere?

Conversely, on a high mountain, where the atmospheric pressure is less than 1 atmosphere, food takes a long time to cook since water boils there at temperatures much lower than 100°C. (The food may not cook at all if the mountain is too high!)

The presence of solutes in the liquid

The boiling point also depends on the *purity* of a substance. If salt is dissolved in water, the presence of ions from the salt means that more energy is required to separate the molecules of water to form vapour. The water will not boil until the average speed of the water molecules increases, i.e. until a higher temperature is reached.

Boiling and evaporation

Boiling and evaporation both involve a change from liquid to vapour. However, you should note these differences:

- Boiling occurs *throughout* a liquid, whereas evaporation takes place only at the surface of the liquid (figure 11.15).
- Boiling takes place at a *definite* temperature, called the boiling point, for a given external pressure, whereas evaporation takes place at all temperatures.
- To keep a liquid at its boiling point, heat energy must be *supplied continuously* to break the bonds holding molecules in the liquid state; the temperature of the liquid stays at the boiling point. Evaporation occurs without the need for any heat to be supplied, and leads to a drop in the temperature of the liquid.

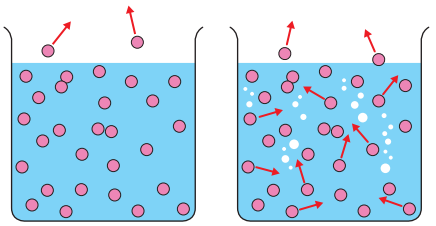


Figure 11.15 Evaporation and boiling.

Thermal expansion and contraction

When a solid or liquid is heated, its particles gain energy and vibrate with greater speeds and amplitudes. Their average distance apart *increases* and this leads to an overall increase in size. In the case of a heated gas, the average speed of the particles increases. This leads to greater average separation distances at constant pressure. Thus, substances usually expand when heated, a phenomenon called **thermal expansion**. When a substance is cooled, its particles move more slowly and the average distance between them decreases. Thus, cooling usually leads to **contraction**.

thermal expansion ›

contraction ›

Demonstration of thermal expansion and contraction

Figure 11.16 shows a metal ball passing easily through a close-fitting metal ring. When the ball is heated, it no longer passes through the ring. This shows that heating causes the ball to expand. If the ball was not heated, but the ring was cooled, the ball would again not pass through the ring. This is because the ring would have contracted when cooled.

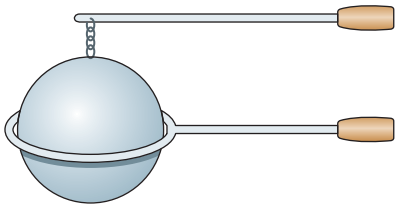


Figure 11.16 The ball-and-ring experiment.

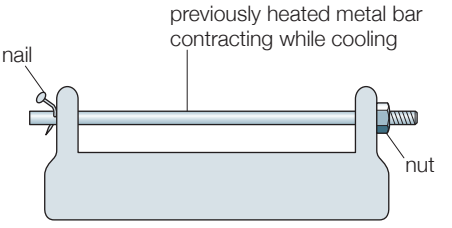


Figure 11.17 Demonstration of the strong force exerted by a metal bar contracting as it cools.

In figure 11.17, a metal bar is heated in a Bunsen flame. A nail is placed in the hole shown straddling a gap and the nut tightened. When the flame is removed, the bar cools and the nail bends. This is because the bar has contracted and is exerting a very large force on the nail.

ITQ9

How do we account for the very large force that the bar in figure 11.17 exerts on the nail?

Some applications involving thermal expansion and contraction

Electricity and telephone wires are usually put in place with sags between the poles. This is to allow for contraction of the wires so that they do not become too tight during cold weather. (We have just seen how great the force of contraction in a metal bar can be.) Bridges are made with gaps between sections, or one of the bridge sections is mounted on rollers, to allow for expansion in hot weather.

Metal railway and cart tyres are heated to make them expand sufficiently to be fitted on to the wheels. As the metal tyres cool, they contract and make a very tight fit on the wheels.

liquid-in-glass thermometer ›

Liquid-in-glass thermometers make use of the fact that liquids expand when heated and contract when cooled.

bimetallic strip ›

Bimetallic strips, used in fire alarms and thermostats, make use of materials that expand by different amounts for the same temperature change. As the temperature rises, a bimetallic strip curves towards the metal that expands less. In the fire alarm shown in figure 11.18, as the bimetallic strip becomes hot, it curves, causing the electrical contracts to close a circuit. A bell or a siren sounds when the circuit is closed.

thermostat ›

The **thermostat** of a refrigerator also depends on a bimetallic strip for its operation. When the interior of a refrigerator gets too warm, the bimetallic strip curves and closes the motor compressor circuit. As the compressor operates, the inside of the refrigerator becomes cool. If the refrigerator gets too cold, the strip curves in the opposite direction and opens the circuit, thus turning off the motor. In this way, the inside of the refrigerator can be kept at a fairly steady temperature.

A jammed metal lid on a jar can be opened by, very carefully, wrapping a cloth soaked in hot water (or running hot water) on the metal top of the jar. The metal screw-top expands, the seal is broken and the jar can be opened.

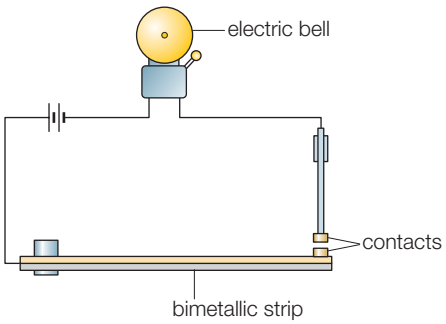


Figure 11.18 Bimetallic strip fire alarm circuit.

Expansion and density

Since a substance generally expands when heated, its volume increases. As the mass remains constant and density = mass/volume, this means that the density decreases when the substance is heated. Thus, hot air is less dense than cold air. Convection currents of air are produced as heated (less dense) air rises and cooler air rushes in to take its place (see chapter 15).

CHAPTER 15

The anomalous expansion of water

Heating does not always lead to expansion. Figure 11.19 shows an arrangement that can be used to demonstrate how the volume of a fixed mass of water varies with temperature. The distance, *h*, can be used as a measure of change in volume of water since *h* is proportional to the volume of water in a uniform tube.

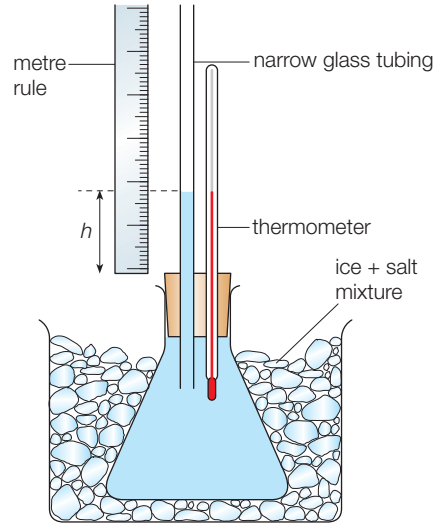


Figure 11.19 Showing how the volume of water changes with temperature.

The graph in figure 11.20 shows that the minimum volume of the water occurs at 4°C. Thus, the maximum density of water occurs at 4°C (since density = $\frac{\text{mass}}{\text{volume}}$ and the volume is smallest at this temperature). The graph shows that water actually expands while being cooled from 4°C to 0°C. Further, at 0°C, the volume of water increases rapidly as the water turns into ice. This

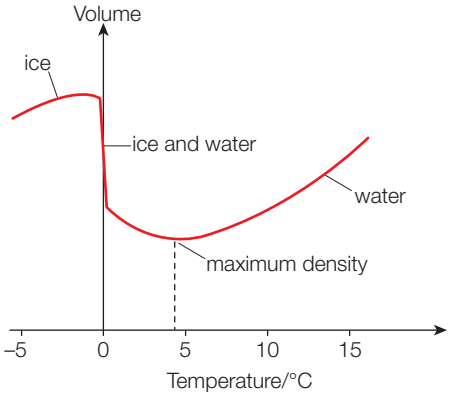


Figure 11.20 Graph illustrating the anomalous expansion of water.

is because at temperatures below 4°C, the water molecules fit together in a different, more open pattern.

anomalous expansion ›

The expansion produced during cooling is described as '**anomalous**' **expansion**, since substances usually contract, rather than expand, when cooled and so this behaviour is seen as strange. In cold countries, the anomalous expansion of water may cause pipes carrying water to burst when the water in them freezes during winter. However, the burst is not discovered until the ice thaws!

The fact that ice has a larger volume than the water from which it is formed also means that ice is *less dense* than water. Thus, ice floats in water (figure 11.21).

ITQ10

Why do cans of soda (soft drinks) burst when left in a freezer too long?

ITQ11

According to the graph in figure 11.20, does water show 'anomalous' behaviour above 4°C?



Figure 11.21 An iceberg floats because ice is less dense than water. Since ice is only slightly less dense than water, we can see the tip of the iceberg when we are above the water!

In this chapter, the kinetic model is discussed in relation to molecular substances.

Figure 11.22 shows that the top of a pond may be frozen even though there is water underneath. As the pond cools from about 10°C to 5°C, the density of the water increases and the cold water sinks to the bottom of the pond. As the temperature drops lower than 4°C, the water becomes less dense and the colder water now rises, leaving water at 4°C at the bottom of the pond. As cooling continues, the temperature at the top drops to 0°C and ice begins to form.

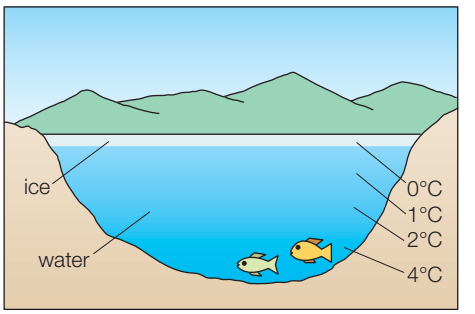


Figure 11.22 Life in a pond during very cold weather.

Since the density of ice is less than that of water, the ice remains on top of the water in the pond. Ice, like water, is a poor conductor of heat. Thus, the water beneath the pond does not get cooled as quickly as before by the cold air above the pond. Hence, it is possible that, for an entire winter season, water can remain in a pond under a sheet of ice formed on the surface. Fish are therefore able to remain alive in the pond throughout the winter by dwelling in this body of water, which is at 4°C approximately.

Chapter summary

- The kinetic model of matter assumes that matter is made up of tiny particles, called molecules. These molecules are always in motion, have strong forces between them and, when they collide, they do so elastically.
- Evidence for the kinetic model comes from the regular shapes and cleavage planes of crystals, Brownian motion, cohesion, adhesion and the elastic behaviour of substances.
- The kinetic model regards solids as being made up of molecules that are very closely packed. At such close spacing, molecules exert strong forces of attraction on each other.
- In liquids, the forces between molecules are not as great, since the average separation distance between molecules in liquids is a little larger than in solids.
- In gases and vapours, the molecules are so far apart that they experience practically no force of attraction and therefore move about fairly independently of one another.
- Energy supplied to a solid (or liquid) can do work against the molecular force of attraction, causing the mean separation distance of molecules to increase. As the average separation distance between molecules increases, the substance may show an overall thermal expansion.
- At certain temperatures and pressures, a substance may show a change of state, such as melting and boiling as it goes from solid to liquid and liquid to vapour, respectively. In the reverse process, the substance may undergo condensation and freezing.

- During a change of state of a pure crystalline substance, the temperature of the substance remains constant. During the process, heat is either released or absorbed depending on the change that is taking place.
- Temperature is explained by the kinetic model as being related to the average speed of the molecules in a substance. An increase in temperature means an increase in average speed.
- Gas pressure and vapour pressure result from bombardment of the internal walls of the container by the molecules of the gas or vapour present within the container.

Answers to ITQs

ITQ1 Yes! During the manufacture of doped semiconductors, atoms of impurities are made to diffuse into solid chips (see pages 474–475).

ITQ2 Tiny molecules of air travelling at high speeds in random directions are probably colliding with the specks of soot, resulting in the jerky movement of the latter.

ITQ3 The drop spreads on the glass surface because the adhesion forces between the glass and water molecules are greater than the forces of cohesion between the water molecules. The drop is spherical on the leaf surface because the force of adhesion between the waxy leaf surface and water is less than the cohesion forces among the water molecules.

ITQ4 **(a)** The graph is horizontal, indicating that there is no change in temperature. This means that the heat being supplied is not being used to increase the kinetic energies of the molecules, but rather to move them apart during the change of state.

(b) This suggests that the 'wax' might not be of one pure material, but might consist of several 'waxes'.

ITQ5 Since only faster particles (i.e. those with greater kinetic energy) are able to leave the surface of the liquid, the average kinetic energy of the molecules remaining in the liquid is reduced – the temperature of the liquid drops, since temperature is a measure of average kinetic energy of molecules.

ITQ6 Unlike water in an earthenware goblet the temperature of the water in the plastic pitcher (see figure 11.13) is not likely to drop by much, since there are no pores in the wall of the container from which water can evaporate and produce a cooling effect. However, if the container is open, some evaporation will take place from the surface and there could be a small temperature fall.

ITQ7 On dry days, the air contains little or no water vapour. The wet clothes lose many more water vapour molecules to the air than they gain from it. On humid days, the wet clothes receive as well as lose water vapour molecules, since both the air and the clothes contain water. Hence the clothes will dry faster on dry days (at the same temperature)

ITQ8 If the external pressure is reduced, a lower vapour pressure within bubbles (corresponding to a lower temperature) can equal the external pressure. Hence boiling can occur at temperatures less than 100°C.

ITQ9 Particles in a solid are very close to each other, and therefore the forces between them are very large. Hence, as the bar contracts it exerts a tremendous force on the nail.

ITQ10 The water in the soft drink expands as ice is being formed from the water in the drink. The large volume of ice formed pushes against the closed can, causing the latter to burst.

ITQ11 No. Water expands as the temperature rises from 4°C. Substances generally expand on heating, so there is no anomaly in this behaviour.

Examination-style questions

1 When smoke is brightly illuminated and viewed through a microscope, small, bright specks of light showing jerky movement are seen. These specks are:

A electrons in the air
B molecules of air
C smoke particles
D photons of light

2 Bridges are built with gaps between sections to allow for:

A contraction during hot weather
B contraction during cold weather
C expansion during hot weather
D expansion during cold weather

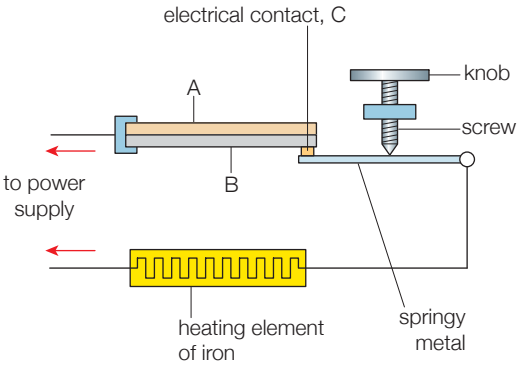
3 Which of the following substances does not contract when cooled from 3°C to 1°C?

A ethyl alcohol B water C copper D steel

4 The conversion of liquid to vapour at a specific temperature and pressure is called:

A boiling B condensation C evaporation D vaporisation

5 The diagram shows a bimetallic strip inside an electric iron. The strip is made of metals A and B.



(i) When the iron is plugged into the power supply and switched on, the heating element gets hot. The bimetallic strip next to the heating element curves and moves away from the contact C. What does this tell you about the relative expansions of the metals A and B on heating?

(ii) Explain what happens to flow of electricity in the heating element of the iron:

(a) when the strip moves away from C.

(b) several minutes after the strip has moved away from C.

(iii) In what direction (up or down) must the screw be adjusted for a 'low' temperature setting on the iron? Explain your answer.

(iv) The bimetallic strip ensures that the iron does not get too hot nor too cold when switched on, that is, it regulates the temperature of the iron. What is the name given to a device, like the bimetallic strip, which helps to keep the temperature of an appliance fairly steady?

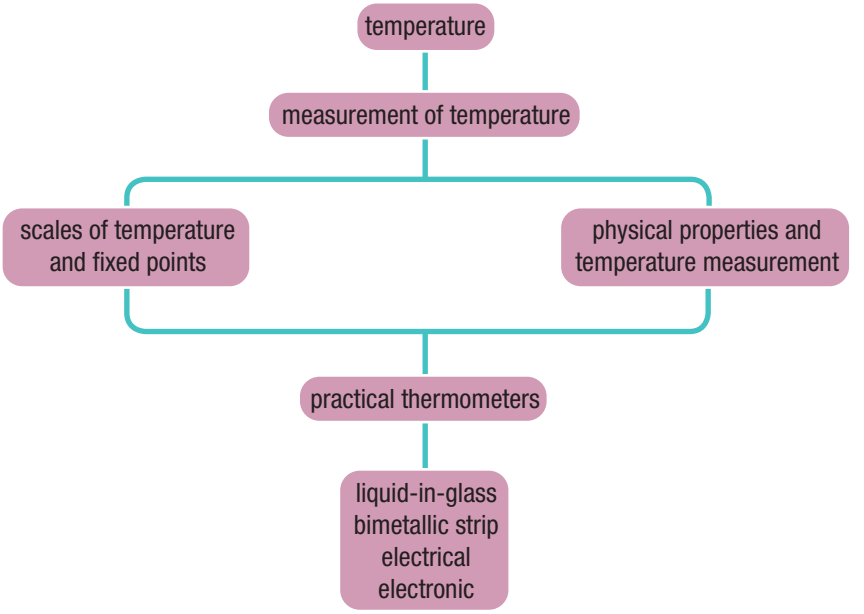
6 The table below summarises a comparison of the properties we associate with solids, liquids and gases. Copy and complete the table.

State	Spacing of particles	Movement of particles	Volume and shape	Special properties
Solid	Closely packed in a regular formation, giving high density	Limited to vibrations about a fixed position		Can be cut, stretched, bent, twisted, and polished. Incompressible
Liquid				Can be poured, forms drops, takes the shape of its container. Incompressible
Gas	Far apart, giving low density	Independent random motion		

12 Temperature and its Measurement

By the end of this chapter you should be able to:

- relate temperature to the direction of net energy transfer
- identify physical properties that vary with temperature and therefore may be used as the basis for measuring temperature
- define the fixed points on the Celsius scale
- relate the use of a thermometer to its design
- use the (approximate) relationship between the Kelvin and Celsius scales: $T = \theta + 273$, where T is the temperature in kelvins and θ is the temperature in °C



Thermometers are everywhere

thermometer

Measurement of temperature is something that is done all the time, and many different kinds of **thermometers** are used to do this.

Clinical thermometers give doctors information about the state of health of a patient. Food-store thermometers monitor the temperature of refrigerated food. Special industrial thermometers are used where chemicals, alloys and semiconductors are manufactured. Information gathered from weather thermometers enables meteorologists to make fairly accurate predictions about the weather. In our homes, we may use oven thermometers and meat thermometers to ensure that the food comes out 'just right'. Wall thermometers that can give both indoor and outdoor temperatures are useful in cold countries, especially in winter – they help people to decide what outdoor clothing they will need. It seems that thermometers are used almost everywhere.

In this chapter, we will distinguish between temperature and heat, explore the measurement of temperature, and discuss the features of some practical thermometers.

What is temperature?

Temperature and hotness

temperature ›

The **temperature** of an object may be considered to be the 'degree of hotness (or coldness)' of the object. A piece of iron that has just been removed from a flame will be 'hotter' than one that has not been in the flame; we say that it is at a high temperature than one that has not been heated.

We learned in chapter 11 (page 172) that the temperature of an object is related to the motion of the particles which make up the object. If the average speed of the particles is large, the temperature of the object is high and the object is hot. If the average speed of the particles is low, the object is cold. Since kinetic energy is related to speed, temperature is a measure of the average kinetic energy of the particles making up a body.

Internal energy and temperature

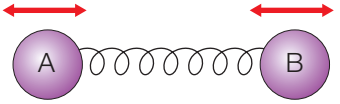


Figure 12.1 Two vibrating masses connected by a spring.

Consider two masses, A and B, lying on a smooth horizontal table and connected by a spring (figure 12.1). If the masses are pulled apart and then released, they vibrate to and fro. As they move apart from each other, the force of attraction between them (caused by the tension in the spring) gets larger. The kinetic energy of the masses becomes converted into potential energy as work is being done in stretching the spring. When the masses momentarily come to a stop, all their kinetic energy has been converted into potential energy. This energy is momentarily stored in the spring. The masses then return towards each other with increasing kinetic energy obtained from the potential energy that was stored in the spring. As the moving masses compress the spring, their kinetic energy is again converted into potential energy and then back to kinetic energy. Thus, in the vibrating system, there is a constant interchange between potential energy and kinetic energy.

ITQ1

What happens to the total energy of such a vibrating system, assuming no energy losses?

internal energy ›
thermal energy ›

Since there are forces of attraction between particles of a substance, a system of moving particles will have both kinetic energy and potential energy. The sum total of the potential and kinetic energies of all the particles making up a substance is called the '**internal energy**' of the substance, or sometimes the '**thermal energy**'. If the substance does not lose energy (e.g. by sharing it with another body with which it may be in contact), the internal energy remains at a constant value all the time.



Figure 12.2 The sparks coming from the grindstone are red-hot.

Experiment shows that a red-hot spark (figure 12.2) which lands on an ice cube melts hardly any of the ice. This shows that even though the spark has a high temperature there is very little energy available to melt the ice. A cup of water at room temperature, however, easily melts an ice cube. This shows that the cup of water (at perhaps 25°C) has more internal energy available to melt the ice than the spark, which was at a much higher temperature (perhaps 1000°C), since the water has many more molecules (though with smaller average kinetic energy) than the spark. So the total thermal energy in a body (e.g. the cup of water), although related to the temperature of that body, is not an indication of the temperature of the body. The spark, although hotter than the water, will have less thermal energy than the cooler cupful of water. In order to assess the temperature of a body, one must know something about the average energy of a molecule of that body.

ITQ2

Which has more internal energy, a drop of boiling water or a cup of water at room temperature?

Temperature and heat energy

heat energy ›

When hot water is mixed with cold water the result is warm water. The cold water gains **heat energy** from the hot water and becomes warmer. The hot water loses heat energy to the cold water and becomes cooler. You feel cold next to an open freezer because heat energy leaves your warm body and travels to the colder ice machine. Thus, heat energy flows from a hotter to a colder object (figure 12.3).

Heat is the thermal energy that flows from a hotter object to a colder object. The temperature of a body determines the direction of the *net* flow of heat energy when a body is brought next to or in contact with another body.

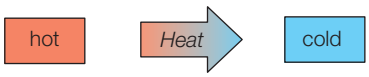


Figure 12.3 The direction of flow of heat energy.

Usually, an object with a tiny mass has little internal energy, even though it is at a high temperature, since such an object would not have a large number of particles. Hence, little energy is available to flow as heat. Thus a spark, falling on to an ice-cube, would hardly have any effect on the temperature or state of the ice-cube, since very little heat energy would have been transferred to the cube. An object with large mass usually has a lot of internal energy, some of which can flow as heat energy, when the object is brought into contact with a colder object. Thus, heat energy from water in a cup at room temperature is large enough to melt all, or a large part, of an ice-cube.

Practical activity 12.1

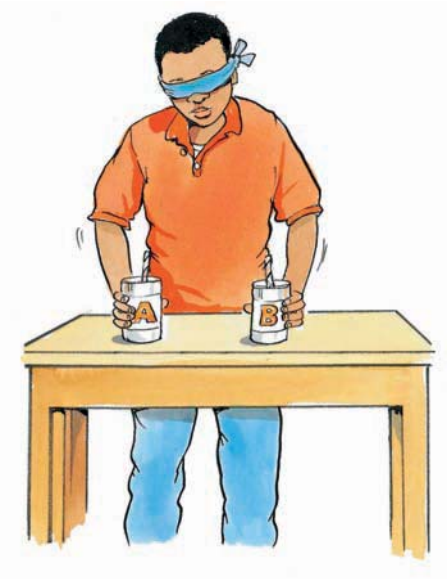


Figure 12.4 Which can is hotter?

Internet search term: thermal sensitivity in animals

Temperature sensitivity of an individual's hands

Your group is to investigate the thermal sensitivity of human hands. (Note: 'sensitivity' refers to the smallest temperature change that can be detected.)

You will need:

- two metal cans of the same size (e.g. calorimeters), labelled A and B
- thermometer graduated in °C (see figure 12.10)
- two stirrers
- water
- ice.

Method

1. Half-fill each can with water at room temperature.
2. Check the temperatures of the water in each can to make sure they are the same.
3. Have someone from your group who is blindfolded place his or her hands around cans A and B simultaneously (figure 12.4). Ask which can feels hotter or colder or whether both seem to have the same temperature.
4. Ask the blindfolded tester to remove his or her hands from the cans.
5. Decrease the temperature of one can by 1°C by adding small pieces of ice.
6. Stir the water in each can just before checking its temperature.
7. Ask the blindfolded tester to repeat step 3.
8. Repeat steps 4 to 7 for a variety of small temperature differences between cans A and B, randomly making A or B colder.
9. Record all data obtained in a suitable manner.

Questions

1. Why is the water stirred just before taking each temperature?
2. How were variables controlled in this experiment?
3. What were some possible sources of experimental error in this activity?
4. What was the smallest temperature difference the blindfolded tester could *consistently* detect? Is this result necessarily the same for each member of the group? How can this be investigated?
5. Why is a metal can preferred to a plastic cup in this experiment?

Extension

Read to find out which parts of the human body are most and least sensitive to temperature changes. Relate the differences in sensitivity to the efficient functioning of the body.

Practical activity 12.2

Temperature sensitivity by gender

Someone says that teenage girls' hands are more thermally sensitive than teenage boys' hands. How could an investigation be designed to test this hypothesis?

Discuss the design of such an investigation critically before submitting it to the teacher (i.e. would the investigation, if carried out as designed, be able to yield a result that can be trusted in relation to the hypothesis? What might be some limitations of the design?).

Measurement of temperature

Why we need an object instrument (the thermometer)

ITQ3

Suggest *two* other reasons why the human body is not a suitable instrument for measuring temperatures.

The human body is able to detect varying degrees of hotness or coldness. So, to what extent can the human body be used for measuring temperature?

When we have a fever, the air around us feels chilly. As soon as the fever breaks, the same air feels warm. Further, two persons might be in the same room, to one, the room is hot, to the other, the room is cool. Which one is correct? We cannot say. The human body cannot therefore give a reliable indication of temperature. A more objective instrument is needed.

thermometer ›

We measure temperature with a **thermometer**. If repeated measurements are taken with a thermometer, the *same* temperature reading is obtained for an object of a given hotness. Further, readings from two properly calibrated thermometers will agree on a measured temperature more closely than estimates from two persons. Thus, a thermometer gives a more reliable indication of hotness than the human body.

To 'calibrate' a measuring instrument is to insert the graduations on its scale, or to check the correctness of the readings it provides.

Temperature scales

Celsius ›
Kelvin ›

A scale of temperature is a 'system' for measuring temperatures. Thus, we find in common use today the **Celsius** system (or the Celsius scale), the Fahrenheit system (or the Fahrenheit scale) and, particularly in science, the **Kelvin** system (or Kelvin scale). Each system for measuring temperatures is based on a **temperature scale**.

temperature scale ›

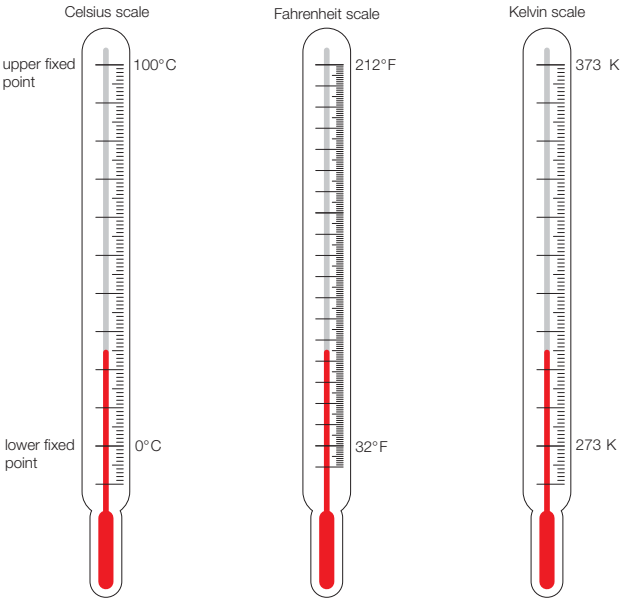


Figure 12.5 Three scales of temperature: Celsius, Fahrenheit and Kelvin.

The Celsius scale

lower fixed point ›

The **lower fixed point** on the Celsius scale is defined as the temperature of pure melting ice at standard pressure. This temperature is defined as 0°C. (Here C is for Celsius.) This lower fixed point is also called the ice point. Figure 12.6 shows how the ice point is determined.

upper fixed point ›

The **upper fixed point** on the Celsius scale is defined as the temperature of steam just above water that is boiling at standard pressure. This temperature is defined as 100°C. This upper fixed point is also called the steam point. Figure 12.7 shows an apparatus used for determining the steam point. The steam outlet to the atmosphere ensures that the steam is at the pressure of the atmosphere.

Celsius, not Centigrade

The Celsius scale is divided up into 100 divisions between the fixed points. Hence, the Celsius scale is a centigrade scale (centi meaning 1/100). A temperature such as 4°C is read as 4 degrees Celsius, not 4 degrees Centigrade.

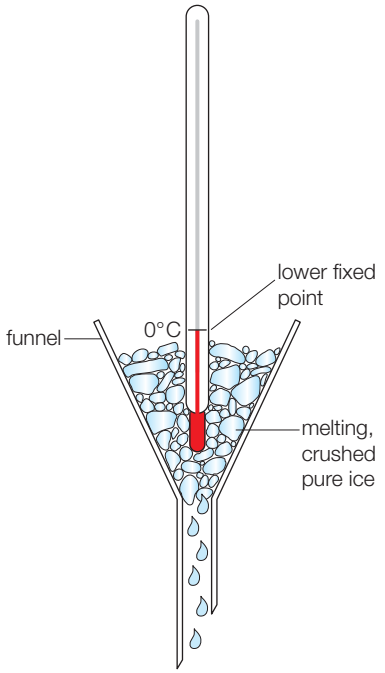


Figure 12.6 Determining the ice point.

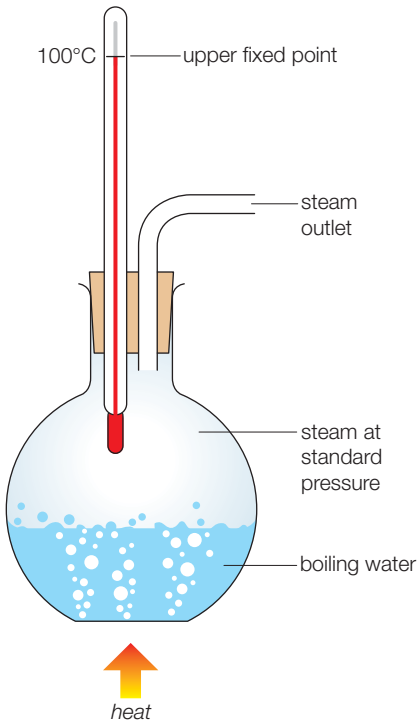


Figure 12.7 Determination of the steam point.

Calibration of the upper and lower fixed points should, ideally, be done when the atmospheric pressure, as read from a lab *barometer*, is standard, i.e. is 1.0 atmosphere.

Standard pressure is equal to a pressure of 1 atmosphere or 1.013×10^5 N m^{-2}.

The temperature interval between the fixed points is divided up into 100 equal divisions (called **degrees**) on the Celsius scale. This scale is therefore also called a **centigrade** scale (from the Latin *centum*, which means 'one hundred', and *gradus*, meaning 'steps').

degree ›
centigrade ›

The Fahrenheit scale

Fahrenheit ›

The **Fahrenheit** scale employs the same fixed points at the Celsius scale, but they are represented by different numbers. The upper fixed point is defined as 212°F and the lower fixed point as 32°F (F for Fahrenheit). There are therefore 180 equal temperature divisions on this scale. Thus a temperature interval of 100 degrees Celsius is equal to an interval of 180 degrees Fahrenheit. The size of the Fahrenheit degree is therefore only 5/9 of that of the Celsius degree.

The Kelvin scale

kelvin absolute zero ›

The **Kelvin** scale assumes that there is an **absolute zero** of temperature. At this temperature, all molecular motion is presumed to cease. Temperature on this scale is specified as kelvin (named after Lord Kelvin, whose original

name was William Thomson. The kelvin, with the symbol K, is the S.I. unit of temperature. Absolute zero is 0 K.

We do not write or say degrees K or °K, but simply K. So we write, for example, 0 K and 273 K.

Capital K is the S.I. symbol for kelvin; small k is the S.I. symbol for kilo (1000).

The size of the degree divisions on the Kelvin scale is the same as that on the Celsius scale. So a change of temperature of one degree Celsius is the same as a change of one kelvin. The fixed point on the Kelvin scale, however, is not the temperature of pure melting ice, but the temperature at which pure water exists in the three states (solid, liquid and vapour) simultaneously at standard pressure. This point is called the **triple point** of water and is defined as 273.15 K. Since a one kelvin division is the same as a one degree Celsius division, the absolute zero of temperature (or 0 K) corresponds to –273.15°C. For calculations in this course, we shall approximate this temperature to –273°C. Thus, the ice point (0°C) corresponds to a temperature of 273 K.

triple point

Conversions between scales

To convert a temperature in Celsius degrees to kelvins we use the equation

$$T = \theta + 273$$

where T and θ represent temperatures in kelvin (K) and degrees Celsius (°C), respectively.

Worked example 12.1: Converting Celsius to kelvin

On a hot day, room temperature is 28°C. What is this temperature in kelvin?

Solution

Celsius temperature, $\theta = 28°C$

Kelvin temperature, $T = ?$

$T = \theta + 273\,K$

$T = 28 + 273\,K$

$T = 301\,K$

Answer: 301 K

To convert between °C and °F we use the equation

$$\frac{\theta}{100} = \frac{F - 32}{180}$$

where F represents temperature in degrees Fahrenheit.

This can be rearranged to

$$\theta = \frac{(F - 32)}{180} \times 100 \text{ and } F = \left(\theta \times \frac{180}{100}\right) + 32$$

Worked example 12.2: Converting Celsius to Fahrenheit

Body temperature is 37.0°C. How much is this in °F?

Solution

Celsius temperature, $\theta = 37.0°C$

Fahrenheit temperature, $F = ?$

$$\frac{\theta}{100} = \frac{F - 32}{180}$$

Therefore

$37.0 \times 180 = 100\,(F - 32)$

$F = 98.6°F$

Answer: 98.6°F

ITQ4

What is 80°F expressed on the Celsius scale?

Practical activity 12.3



Demonstrating fixed points on a simple electrical thermometer

Caution: Hot surface

The topic of electricity is treated in section D. However, this demonstration is simple enough for you to put together at this stage, with a little assistance from your teacher. You must consult with your teacher concerning precautions to be taken, the method of connecting the circuit, the method of using the instruments and the measurement units to be used.

You will need:

- bead-type thermistor
- silicone sealant
- two insulated connecting wires with crocodile clips
- beaker with boiling water
- beaker with melting ice
- multimeter.

(The connection to the multimeter must be done by the teacher. An alternative circuit using a milliammeter is shown in figure 12.8, in case a multimeter is not available.)

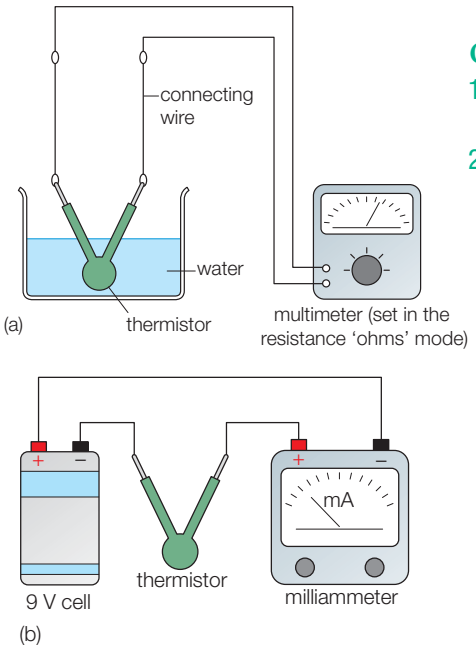


Figure 12.8 (a) Upper fixed point on a simple electrical thermometer. (b) An alternative circuit: the thermistor is connected to a milliammeter (or to a multimeter set in the current 'milliamperes', mA, mode).

Method

1. Waterproof the thermistor bead and legs with silicone sealant a few days before the activity and allow it to harden. Have the teacher inspect the waterproofing.
2. Switch the multimeter to the 'ohms' scale and zero the meter.
3. With the switch in the 'off' position, connect the thermistor to the multimeter (under the supervision of your teacher) as shown in figure 12.8 (a). Then turn the meter back to the 'ohms' scale of step 2.
4. Insert the thermistor into boiling water, taking care to submerge only the waterproofed portion, and note the 'ohms' reading, R_1, on the multimeter or the 'mA' reading, I_1, on the milliammeter. (The ohm is the S.I. unit for electrical 'resistance'; the ampere is the S.I. unit for electric current.)
5. Insert the waterproofed portion of the thermistor in a beaker containing crushed melting ice and note the 'ohms' reading, R_2, on the multimeter, or the 'mA' reading, I_1, on the milliammeter.

Questions

1. What temperatures correspond to the readings R_1 and R_2 (or I_1 and I_2)?
2. Calibrate the temperature scale of the meter by plotting a graph of resistance (or current) against temperature from 0°C to 100°C.

   (Since, in calibrating a thermometer scale, the temperature interval between the fixed points is divided up into 100 equal temperature divisions, we are defining that the property (e.g. resistance) varies linearly with temperature. When this relationship occurs, the graph of the property vs temperature is a straight line. We therefore need only the two fixed points in order to obtain this straight line.)
3. Note the value of the resistance (or current) at room temperature. Using this value, determine the room temperature from your graph.
4. How does this value compare with the room temperature taken on a

MATHEMATICS: linear graphs

Internet search terms: thermistor thermometer

laboratory liquid-in-glass thermometer? Comment on your answer.

Extension

Read to find out whether equal changes in temperature result in equal changes of resistance in a thermistor. Also, find out which kind of thermometer is assumed to have a linear scale (i.e. equal changes on the thermometer represent equal changes in temperature).

Linear variation

If quantity A varies linearly with quantity B, then any change that takes place in B will be proportional to the change in A that caused it. So, in a mercury-in-glass thermometer, the change in the volume of the mercury is proportional to the change in temperature. A graph of quantity A against quantity B will be a straight line.

linear relationship ›

liquid-in-glass thermometer ›

MATHEMATICS: linear relationships

Electrical thermometers

One advantage of thermometers based on electrical properties (such as resistance) is that the actual reading can be done elsewhere. The readout display may be linked to the sensing element by wires or even by wireless transmission. So, for example, an outside temperature could be easily read on a display indoors.

sensitivity ›

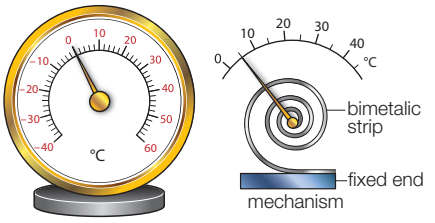


Figure 12.9 A bimetallic strip thermometer and its mechanism.

Thermometers in practice

Physical properties

A practical thermometer is made using a substance which has a physical property that varies *linearly* with temperature. The constant-volume gas thermometer, in which the pressure of the gas varies nearly linearly with temperature, is thought to come closest to this requirement. In a **liquid-in-glass thermometer**, the volume of the liquid changes with temperature in a fairly linear way. The bimetallic strip thermometer (figure 12.9) makes use of the fact that one of the metals expands more than the other for a given change in temperature. The greater the temperature change, the greater is the curvature of the strip, and the relationship between curvature and temperature is also almost linear. Over a small range, the electrical resistance of a conductor or semiconductor also changes with temperature in a fairly linear fashion.

All of these physical properties, and others, have been used to make thermometers. The various types of thermometers have different advantages and disadvantages, depending on the property used to measure temperature.

Required features of thermometers

The features required in a particular thermometer depend on how it is to be used. Sometimes, for example, **sensitivity** (the ability to detect tiny changes in temperature) is important. In other situations, it is the speed with which the thermometer responds that matters most. The temperature range over which a thermometer can operate is also important. The type of thermometer chosen, too, should be convenient for use in a particular application.

The discussion of liquid-in-glass thermometers (table 12.1), and the comparisons of these with other types of thermometers (table 12.2), shows how these factors are important.

Liquid-in-glass thermometers

Liquid-in-glass thermometers generally have mercury or alcohol as the thermometric fluid. Table 12.1 shows a comparison between mercury and ethyl alcohol, and water is included to show why it is not a good choice of liquid.

Table 12.1 Advantages and disadvantages of mercury and ethyl alcohol for use in liquid-in-glass thermometers. Water is not a good choice for such a thermometer.

	Mercury	Ethyl alcohol	Water
1	Boils at 357°C	Boils at 79°C	Boils at 100°C
2	Freezes at –39°C	Freezes at –117°C	Freezes at 0°C
3	Silvery appearance, fairly easily seen	Can be coloured and made easily seen	Can be coloured to make it easily seen
4	Expensive	Cheap	Cheap
5	Expands uniformly with temperature	Expands fairly uniformly with temperature	Does not expand uniformly with temperature
6	Vapour is toxic	Vapour is non-toxic	Vapour is non-toxic

Alcohol thermometers can be used for measuring daily temperatures since these temperatures are well within the range between its boiling and freezing points. In addition, since alcohol is cheap, these thermometers are easily affordable.

IT Q5
Which property of water makes it a poor choice for use in a thermometer?

IT Q6
Would mercury or alcohol be more suitable for use in thermometers measuring the sort of temperatures found (i) in Arctic regions and (ii) when vegetables are boiled?

IT Q7
Would mercury or alcohol be more suitable for thermometers that are to be used where there are children around? Give a reason for your answer.

A liquid-in-glass thermometer consists of three main parts: a hollow bulb to serve as a reservoir for the liquid, a narrow **bore** in which the liquid expands, and a stem on which the scale markings are made. The glass hardly expands at all with changing temperature.

Laboratory thermometers

In high school laboratory experiments, temperatures generally range from –10°C to 110°C. The mercury thermometer, a liquid-in-glass thermometer (figure 12.10), is quite suitable for such a range of temperatures since mercury boils at about 360°C and freezes at –39°C. The silvery mercury thread in the bore is easily seen, and this makes it easy to read the temperature. The marked divisions are uniformly spaced on the stem. The uniform expansion of mercury with temperature implies that equal changes in temperature correspond to equal changes in distances between marked divisions. Since the bore is narrow and the bulb containing the mercury is large, degree changes in temperature result in an easily visible change in length of the mercury thread. This means that the thermometer is *sensitive*. The glass of the bulb of the thermometer is very much thinner than the glass of the stem, so that heat is conducted more quickly through it.

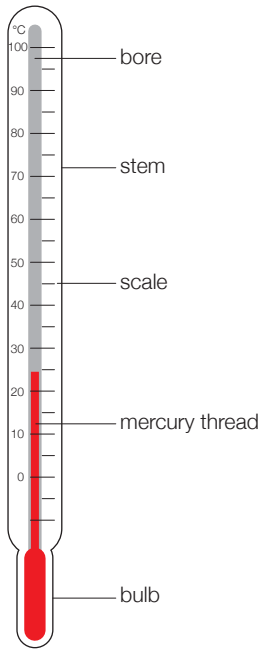


Figure 12.10 Liquid-in-glass laboratory thermometer.

Clinical thermometers

Clinical thermometers (figure 12.11) are used for measuring body temperatures. Table 12.2 summarises some features of liquid-in-glass and electronic clinical thermometers, for comparison.

ITQ8
Why is the bulb of the clinical liquid-in-glass thermometer made of thinner glass than the stem?

ITQ9
What are *two* possible dangers to be guarded against when using a mercury-in-glass clinical thermometer?

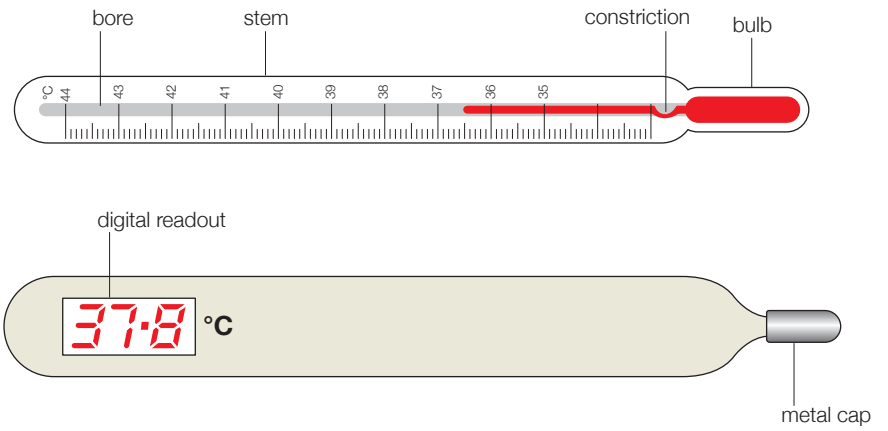


Figure 12.11 Liquid-in-glass and electronic clinical thermometers.

Table 12.2 Comparison of a liquid-in-glass and an electronic clinical thermometer.

	Feature	Liquid-in-glass	Electronic (semiconductor)
1	Speed of response	(i) The wall of the bulb is made very thin so that heat is conducted very quickly to the liquid inside. (ii) However, the volume of liquid is large, so it takes some time for the liquid in the bulb to attain the temperature of the body.	(i) Semiconductors respond very quickly to temperature changes. (ii) The wall covering the semiconductor element is made of metal so as to conduct heat very quickly. (iii) The semiconductor and metal covering are very small. Thus they quickly attain the temperature of the body, and do not change that temperature much.
2	Sensitivity	(i) The volume of the bulb (and liquid) is large. Therefore a small change in the temperature being measured results in an appreciable change in the volume of the liquid. (ii) The bore of the stem is so fine that a small change in volume of liquid shows up as a large change in the length of the thread.	(i) The electrical resistance of semiconductors changes appreciably with temperature. (ii) An electronic circuit further amplifies the change. Hence, semi-conductor thermometers can be made quite sensitive.
3	Ease of reading	The wall of the stem is curved and behaves like a lens. It magnifies the view of the thread of liquid in the narrow capillary bore, making the thread easily visible.	The thermometer has a digital readout, so the temperature is easily read.
4	Retention of the highest reading	When the liquid expands beyond the constriction, it cannot return to the bulb until the thermometer is shaken. Thus, the thermometer can be read conveniently even after removal from the patient.	The thermometer 'beeps' when it has detected the highest reading. The thermometer also has an electronic memory which retains the highest reading obtained.
5	Convenience of handling	Short length. (Only a small temperature range of about 35°C to 42°C is needed, in any case.)	Short length. Electronic circuits and batteries are easily miniaturised.

Weather thermometers

Some weather applications make use of liquid-in-glass thermometers. Figure 12.12 shows a maximum-minimum thermometer.

Since alcohol expands much more than mercury, at high temperatures the steel index in the right tube will be pushed upwards, indicating the highest temperature reached. Note the space on the top of the right tube to allow for the alcohol to be pushed upwards.

At lower temperatures, the alcohol contracts much more than the mercury. The latter now pushes a steel index up the left tube to indicate the lowest temperature reached.

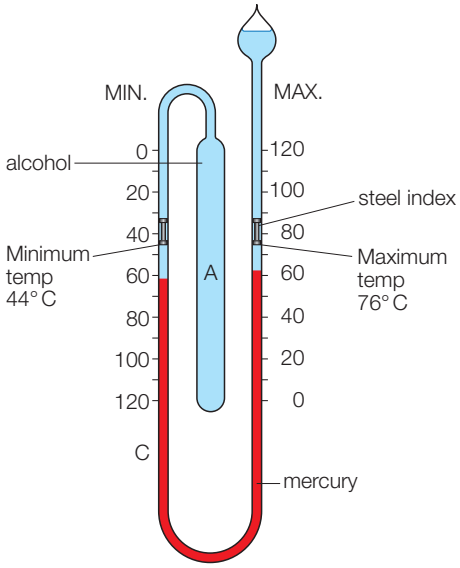


Figure 12.12 Maximum-minimum thermometer.

Thus, this type of thermometer retains the highest and lowest temperatures of the day and can be read at a convenient time. A magnet is used to reset the steel indices.

Industrial (thermocouple) thermometer

A thermocouple thermometer is used in some situations involving high temperatures (e.g. in measuring the temperature of a furnace). Figure 12.13 shows the principle of operation of such a thermometer.

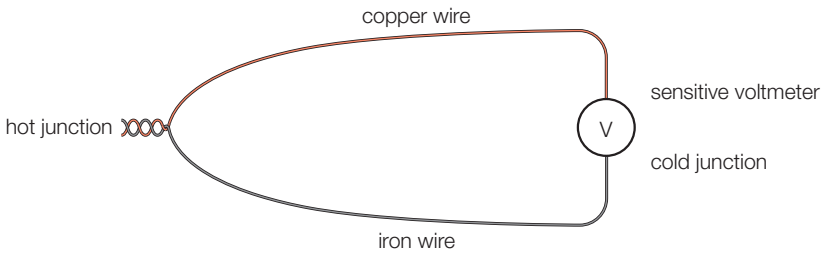


Figure 12.13 Principle of operation of a thermocouple.

Wires of two dissimilar metals are joined together at one end, called the 'hot junction'. The other 'cold' ends are connected to a very sensitive voltmeter which measures the voltage generated when there is a temperature difference between the hot and cold ends. This voltage varies with temperature and hence the voltmeter could be calibrated to read temperature. In use, the radiation from a furnace, for example, is directed to the hot end and the voltage is read and converted to temperature.

The main advantage of a thermocouple thermometer is that it can be used to measure extremely high temperatures. One disadvantage, however, is that the voltage generated by the thermocouple does not vary linearly with temperature. Infrared thermometers have been developed which can measure furnace temperatures.

Internet search terms: infrared thermometer

Chapter summary

- Temperature is a measure of the degree of hotness or coldness of a body. In terms of the Kinetic Theory, temperature is a measure of the average energy of the particles that make up a substance.
- The internal energy (or thermal energy) of a substance is the sum total of the potential and kinetic energies of the particles that make up the substance.
- The temperature of a substance determines the direction of the net energy transfer, called heat energy, between two objects. The net flow of heat energy takes place from a hotter object to a colder object.
- Physical properties that vary with temperature are used for measuring temperature. Properties such as gas pressure, expansion of liquids and solids, and electrical resistance of conductors and semiconductors have been utilised in making thermometers.
- Some factors considered in the design of thermometers, and the choice of thermometers for particular purposes, include speed of action, sensitivity, range of temperature and convenience of use.
- The fixed points on the Celsius scale are the steam point (temperature of steam above boiling water at standard pressure) and the ice point (the temperature of pure melting ice at standard pressure).
- The Kelvin scale assumes an absolute zero of temperature, which occurs at approximately –273°C. Temperature T on the Kelvin scale is related to temperature θ on the Celsius scale by the formula: $T = \theta + 273$.

Answers to ITQs

ITQ1 The total energy remains constant.

ITQ2 The cup of water at room temperature (since it can melt much more ice than the drop of boiling water).

ITQ3 The range of temperatures to be measured is limited since temperatures above or below body temperature can cause damage to the skin. Also, the skin is not very sensitive to slight changes in temperature.

ITQ4 27°C

ITQ5 Water does not expand uniformly with temperature.

ITQ6 **(i)** Alcohol; **(ii)** mercury.

ITQ7 Alcohol. If the thermometer breaks, escaping alcohol vapour is less toxic than mercury vapour.

ITQ8 To allow heat to be conducted very quickly to the liquid contained in the bulb.

ITQ9 The broken glass can cut the skin; the spilt mercury gives off a toxic vapour.

Examination-style questions

1 Temperature is a measure of:

A heat
B hotness
C internal energy
D potential energy

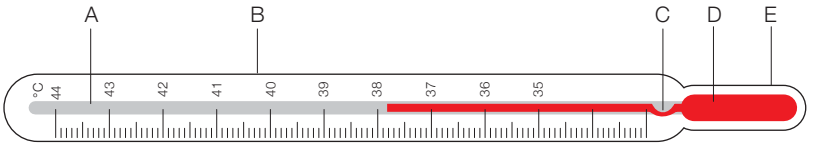
2 What kind of thermometer is BEST suited for an oven?

A alcohol
B bimetallic strip
C mercury
D water

3 An object has a high temperature when:

A the molecules get hot
B the molecules move at high speeds
C the molecules move at slow speeds
D the molecules are stationary

4 The diagram below shows a special type of thermometer.



(i) What is the particular name given to this type of thermometer?

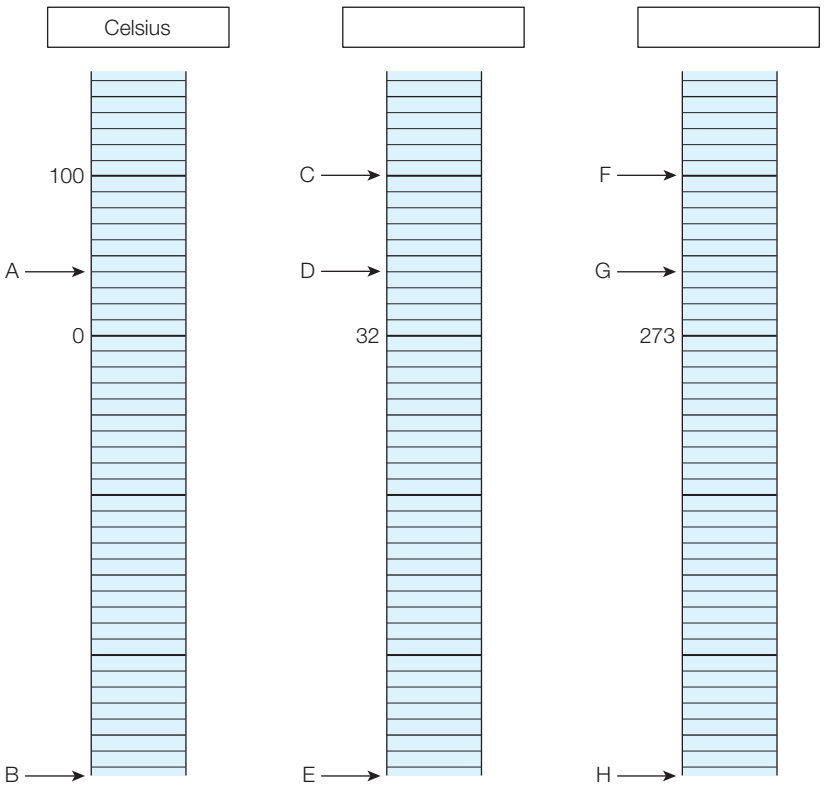
(ii) State the reading on the thermometer.

(iii) Label the parts A to E.

(iv) State the purpose of the part labelled C.

(v) Why is the part labelled E made of very thin glass?

5 The diagram on the left shows three scales of temperature.



(i) Name the temperature scales (one is done for you).

(ii) Indicate the temperatures shown at the arrows on these scales.

6 It is quite common in Caribbean countries for hot liquids (like tea or coffee) to be cooled quickly *either by*:

(i) pouring the liquid from one cup, A, into another, B, and then from B back into A, the two cups being separated from each other as the pouring proceeds,
or by

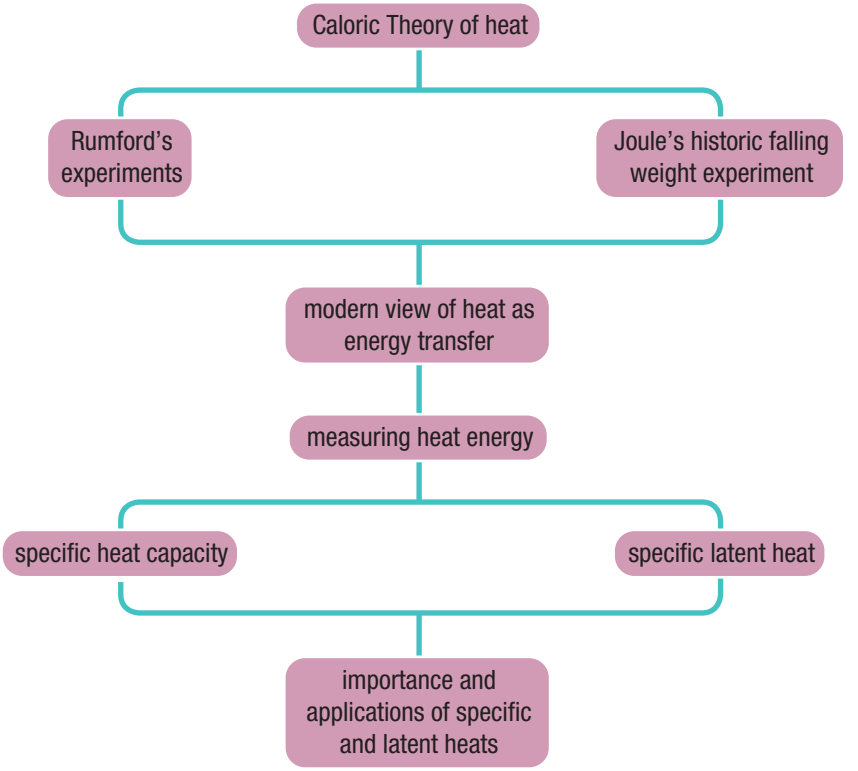
(ii) blowing over the hot liquid (if it is tea or coffee, say) before sipping it, if it is to be drunk.

Explain, in terms of the Kinetic Theory, why both methods are very effective in cooling the liquids.

13 Measuring Heat Energy

By the end of this chapter you should be able to:

- differentiate between the Caloric Theory and the Kinetic Theory of heat as they existed in the 18th century
- explain how Rumford's cannon-boring experiments provided evidence against the Caloric Theory
- recall the role of Joule's experiment in establishing the principle of conservation of energy
- define specific heat capacity
- define heat capacity
- recall the equation $E_H = mc\Delta T$ and use it to solve problems on specific heat capacity and heat capacity
- describe experiments to determine by electrical heating the specific heat capacity of metals and liquids
- perform experiments to measure specific heat capacity by the method of mixtures
- define specific latent heat
- recall the equation $E_H = ml$ and use it to solve problems on specific latent heat
- perform an experiment to determine the specific latent heat of fusion of ice using a container of negligible heat capacity
- describe an experiment to determine the specific latent heat of vaporisation of water



The Caloric Theory of heat

caloric ›

In the early 18th century, scientists thought that heat was an invisible, fluid substance. They even had a name for this substance – '**caloric**'. They thought that all matter possessed caloric. When an object became warm, it was because more caloric had flowed into it. When, conversely, a substance cooled, it was because it had lost caloric. The unit of heat was taken as the calorie, where a calorie of heat (abbreviated to cal) was defined as follows:

A kilocalorie (kcal) is 1000 calories. The kilocalorie, abbreviated to Cal (with a capital C), is the unit used when we speak of food 'calories' (figure 13.2).

calorie ›

One calorie is the quantity of heat which raises the temperature of one gram of water by one Celsius degree.

A problem with the caloric model of heat was that the mass of this 'caloric' fluid could not be detected. An object weighed before and after gaining caloric still had the same mass.

This chapter will discuss experiments that led to the abandonment of the Caloric Theory of heat and its replacement by an energy interpretation of heat. Some of the ways in which we measure heat energy are then described.



Figure 13.1 An iron pot and an aluminium pot of equal mass. In which pot is an egg likely to boil faster, starting from room temperature? (See answer to ITQ11, page 211.)

Cannons, as used in the 18th century, were very large war guns that fired heavy metal balls as the 'bullets' to destroy ships (figure 13.3). The barrel of a cannon was made by boring a solid brass cylinder.



Figure 13.3 A cannon at The Garrison, Barbados, near CXC headquarters.

Nutrition Facts

Servings size 1/4 of Package (20 g)
Servings Per Container 4

Amount per serving	
Calories 90	Calories from fat 20
	% Daily Values*
Total fat 2.5 g	4%
Saturated Fat 1.5 g	8%
Cholesterol 0 mg	0%
Sodium 30 mg	1%
Total carbohydrate 14 g	5%
Dietary Fibre 2 g	7%
Sugars 12 g	
Protein 0 g	

Vitamin A 0% * Vitamin C 0%
Calcium 2% * Iron 8%

*Percent Daily Values are based on a 2,000 calorie diet.

Figure 13.2 We can find the term 'Calories' (kcal) on food labels.

Experiments on the nature of heat

Count Rumford's experiments

The American-born Benjamin Thompson, later known as Count Rumford, was puzzled that a brass cannon tube and the chips that came from it became very hot when the hole was being bored. The chips caused water to boil! Rumford was acquainted with the prevailing Caloric Theory, in which heat was thought of as a massless fluid that could enter or leave objects. Where did the caloric (heat) come from when the cannon was being bored? It seemed that the supply of caloric was inexhaustible; as long as the boring of the cannon continued, the brass chips (and the cannon tube) felt very hot to the touch – i.e. gave off caloric.

Thompson decided to test a theory that maybe caloric (heat), which was already present in substances, was being *squeezed out* of the brass during the boring. Thompson used equal masses of brass chips and unbored brass at 99°C and placed them in equal amounts of water at 15°C. Both types of brass gave the same temperature rise in the water. Hence, they both gave off the same amount of caloric and therefore must have had the same caloric content – a contradiction of the 'squeezing out' theory.

Thompson also measured the masses of objects before and after heating and found that the masses before heating were the same as the masses after heating. Thus, if the fluid, caloric, existed, it had no mass. To Thompson, the concept of a massless fluid was unsatisfactory. His experiments, carried out in the late 18th century, contributed significantly to the abandonment of the Caloric Theory.

Thompson suggested that heat was not a substance as such, but was obtained *from friction*. That was why the supply of heat was almost inexhaustible; as long as the boring continued, heat was produced. Thompson

also suggested that the motion involved in the boring caused 'particles' within the substance to *vibrate*. He suggested that heat flow was related to the *vibrating motion of these particles*.

Joule's historic experiment

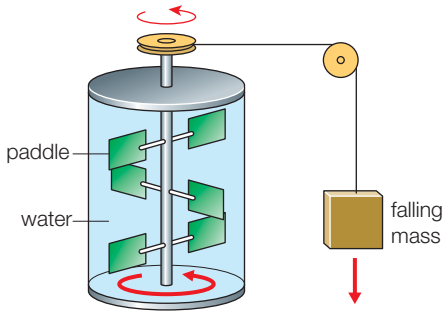


Figure 13.4 Joule's falling weight experiment.

An English brewer, James Prescott Joule, made some measurements which confirmed Rumford's conclusions. In one of Joule's experiments carried out in 1843 (figure 13.4), a mass, *m*, falling a distance, *h*, caused a paddle to turn in a container of water. The work done in overcoming the friction forces between the paddle, and the water produced heat, which caused the temperature of the water to rise. Joule calculated the work, *W*, done by the falling mass ($W = mgh$; see chapter 9, page 130) and the number of calories of heat produced. Joule's calculations showed that a fixed amount of work done resulted in 1 cal of heat being produced. (The fixed amount of work has been established as 4.186 joules, or approximately 4.2 J.) Thus

4.186 J = 1 cal

mechanical equivalent of heat ›

The equivalence between work and heat was now established, and the work that had to be done to produce 1 calorie of heat became known as the '**mechanical equivalent of heat**'. Thus, the heat gain in Joule's experiment was the result of mechanical work being done. Since work is an energy conversion process, it meant that the potential energy of the mass must have been converted to heat energy by the mechanical work process. Other experiments using *electrical* energy, carried out by later researchers and by Joule himself, produced the same numerical results initially obtained by Joule. The Caloric Theory of heat was therefore eventually abandoned in favour of the **energy transfer theory** that we use today.

energy transfer theory ›

Internet search terms: Joule's experiment/ mechanical equivalent of heat

Since 1 cal of heat was 'the quantity of heat which raises the temperature of one gram of water by one degree Celsius' (see above), this meant that 4.186 J of energy transferred to 1 g of water would raise the temperature of the water by 1°C. Similarly, a drop in the temperature of an object came to be regarded as being caused by transfer of heat energy away from the object. (The direction in which heat transfer takes place is shown below in figure 13.6.)



Figure 13.5 How far must Leon walk to work off the energy from a piece of chocolate cake?

Worked example 13.1: Working off excess calories

Leon loves chocolate cake.

(a) How much work must Leon do to burn up 500 cal from a piece of chocolate cake? (Assume that all the energy provided by the chocolate cake is converted to mechanical work.)

(b) If Leon exerts an average horizontal force of 400 N during walking, how far must he walk to 'work off' the calories?

Solution

(a) 500 Cal = 500 × 1000 cal
= 500 000 cal
since 1 cal = 4.2 J
work needed = 500 000 × 4.2
= 2 100 000 J
Answer: 2.1 MJ

CHAPTER 9

(b) Recall from chapter 9, page 128 that

work = force × distance travelled

Let distance walked in the direction of the force = d

$2\,100\,000 = 400 \times d$

$d = 5300$ m (2 significant figures)

Leon must walk approximately 5 km! Actually he might need to walk only about 2 km, since the body is only about 30% efficient in converting energy from food into work. The body gets hot in the conversion process and heat is lost to the atmosphere. Some of the energy from the cake is also used to keep the body metabolism going.

BIOLOGY: metabolism (respiration)

Heat as energy transfer

Heat and internal energy

CHAPTERS 11 & 12

We can explain the concept of heat energy using the Kinetic Theory of matter (chapter 11, page 169). According to this theory, particles of matter are always in constant motion. Thus they possess kinetic energy. The particles also attract and repel each other, so they also possess potential energy. (See chapter 12, where this is discussed in more detail.) The total energy possessed by all the particles of a substance is called the **internal energy** of the substance (page 184). A substance of large mass would therefore tend to possess more internal energy than one with a smaller mass.

internal energy

The temperature of a substance is, however, a measure of the average kinetic energy of the particles. Thus, for example, a tiny red-hot spark can have a high temperature but little internal energy (see page 184).

Heat represents internal energy that flows from an object at a higher temperature to another object at a lower temperature. The average kinetic energy of the molecules of the hotter object decreases while that of the colder object increases (figure 13.6). Thus, there is a flow of energy from an object at a higher temperature to one at a lower temperature. A hot solid that is in contact with a cold solid becomes colder because there is a net flow of **heat energy** from the hotter body to the colder one. (Ways in which heat energy is transferred are discussed in chapter 14.)

heat energy

CHAPTER 14

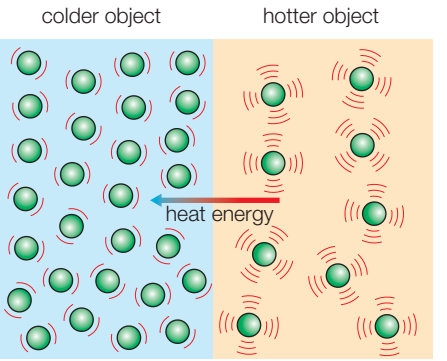


Figure 13.6 The direction of heat energy transfer when two solids at different temperatures are brought into thermal contact.

Specific heat capacity

Heat energy and specific heat capacity

In physics, the word 'specific' is used to refer to unit mass.

Unless there is a change of state, the temperature of a substance rises when it gains heat energy. Temperature rise is therefore used in determining how it gains heat energy. Temperature rise is therefore used in determining heat energy transfer. Experiment shows that different amounts of heat are needed to produce the same rise in temperature in similar masses of different materials.

specific heat capacity ›

The specific heat capacity, c, of a substance is the number of joules of heat energy needed to raise the temperature of 1 kg of the substance by 1 kelvin.

Table 13.1 gives examples of the specific heat capacity of various materials.

Thus, if we want to raise the temperature of 1 kg of a substance by 1 K, c joules of heat energy must be supplied to it. Then, to raise the temperature of a mass, m, by ΔT kelvins, the heat needed will be $mc\Delta T$. If we represent this heat by E_H, we have

$$E_H = mc\Delta T$$

The Greek symbol Δ ('delta') is used to mean '*change in*'. So ΔT means change in T, that is:

ΔT = final temperature
– initial temperature

Table 13.1 Specific heat capacities of some solid and liquids

Substance	Specific heat capacity, c/J K^{-1} kg^{-1}
aluminium	900
copper	390
iron (or steel)	450
wood	1700
ethanol	2400
water	4200
ice	2100
glass	840

ITQ1

How many joules of heat energy are required to raise the temperature of 5.0 kg of water from a temperature of 25°C to 35°C? (Hint: use the equation $E_H = mc\Delta\theta$, and refer to table 13.1 for the value of the specific heat capacity of water.)

ITQ2

Which of the substances in table 13.1 would undergo the smallest temperature rise if the same quantity of heat energy was transferred to equal masses of each substance?

A change in temperature of one Celsius degree is the same as a change of one kelvin, so we can still use the equation $E_H = mc\Delta T$ and $E_H = mc\Delta\theta$ interchangeably since $\Delta\theta$ is equal to ΔT, numerically.

Once the values of m, c and ΔT are known, the heat energy, E_H, transferred can be determined.

Specific heat capacity units

Although one kelvin is equal to one Celsius degree, the use of the S.I. unit, J K^{-1} kg^{-1}, rather than J(C°)$^{-1}$ kg^{-1}, is preferred when writing the unit symbol for specific heat capacity. The solidus representation of the unit (J/C° kg) is still in use in some textbooks. For convenience, the unit J K^{-1} g^{-1} is also sometimes used. Note that the unit for Celsius *temperature* is °C, but the unit for Celsius temperature *change* is C°.

It can easily be shown that 1 J K^{-1} g^{-1} is the same as 1 kJ K^{-1} kg^{-1} or 1000 J K^{-1} kg^{-1}.

Heat capacity and specific heat capacity

heat capacity ›

The heat capacity, C, of an object is the number of joules of heat energy required to raise the temperature of the object by 1 K.

Thus

$$C = \frac{E_H}{\Delta T} \text{ or } E_H = C\Delta T$$

where E_H is the energy supplied to the object and ΔT is the resulting temperature rise.

It follows, from equations $E_H = mc\Delta T$ and $E_H = C\Delta T$, that for an object of mass m kg made of a substance of specific heat capacity c, then

heat capacity of the object, $C = mc$

ITQ3

What is the heat capacity of an aluminium can of mass 200 g?

Measuring the specific heat capacity of a liquid by an electrical method

Figure 13.7 illustrates an electrical method for determining the specific heat capacity, c, of a liquid.

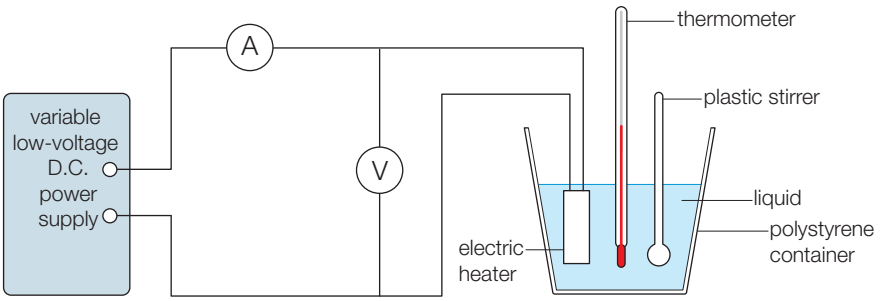


Figure 13.7 Apparatus for determining the specific heat capacity of a liquid by an electrical method.

An electric current is passed for a given time through a heating coil immersed in the liquid. As current passes through the coil, electrical energy is converted into thermal energy and the coil becomes, initially, hotter than the liquid. Heat energy, E_H, then flows from the hotter coil and causes the temperature of the cooler liquid and container (called a **calorimeter**) to rise.

If the potential difference (or voltage), V, across the coil and the current, I, through the coil are kept *steady* for the given time, t, then the energy supplied by the current is given by $E_H = IVt$ (see chapter 26, page 412). The temperature of the liquid and container rises by an amount $\Delta\theta$. If the liquid has mass m and specific heat capacity c, and the container has mass m_1, and specific heat capacity c_1, then the energy supplied to them is

$$E_H = mc\Delta\theta + m_1c_1\Delta\theta$$

Thus we get the following equation:

$$IVt = mc\Delta\theta + m_1c_1\Delta\theta$$

The specific heat capacity, c, of the liquid is easily calculated if all the other quantities in the equation are known.

Note some of the precautions to be taken in an experiment of this nature:

1. The power supply is adjusted to keep the current at a steady value throughout the time, t.
2. Since hotter parts of liquids tend to move up over cooler portions, the liquid must be stirred gently to mix it well just before its temperature is taken. This is done using a very light plastic stirrer (i.e. a stirrer that has a small heat capacity and therefore would not take up much of the heat energy that was supplied by the coil).
3. Because this experiment is of relatively long duration, heat energy could be lost to (or gained from) the surroundings. A method known as **Rumford's cooling correction** is used to minimise heat loss to the atmosphere. The liquid is first cooled to about 10°C below room temperature. The heater is turned on until the temperature of the liquid rises to 10°C above room temperature. In the first part of the heating, the liquid gains heat from the atmosphere because it is at a lower temperature than the atmosphere. In the second part of the heating, the liquid loses heat to the atmosphere because the liquid is hotter than the atmosphere. In this way, the overall heat lost to and gained from the atmosphere in the experiment should be approximately zero.

Instead of an ammeter and voltmeter, a joulemeter (figure 13.8) can be used to directly measure the energy supplied.

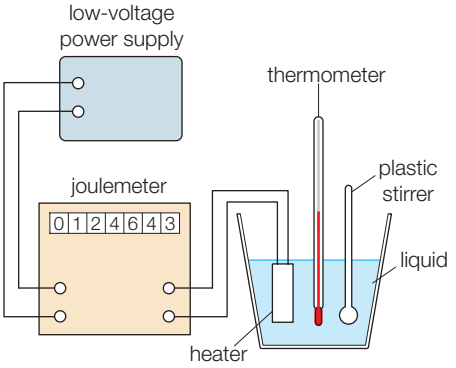


Figure 13.8 The joulemeter gives a direct reading for the energy, E_H, supplied by the electrical heater.

calorimeter ›

CHAPTER 26

Using a Styrofoam container

If, for the container, m_1 is very small and its specific heat capacity is also negligible, we can ignore the term $m_1c_1\Delta\theta$ in the equation. Nowadays we use a Styrofoam (polystyrene) cup as a 'calorimeter'. Since this material is a very poor conductor of heat, it hardly warms up at all. Styrofoam also has a very low density and so the mass of the cup is small. Its heat capacity (m_1c_1) is therefore very small indeed and so may be ignored.

Rumford's cooling correction ›

4 When the current is switched off, the highest temperature reached should be taken to allow for the heat from the coil assembly to be transferred to the water.

5 If the vessel containing the liquid is a good conductor of heat, it should be surrounded by an insulating jacket (see figure 13.9) to minimise heat transfer by conduction to or from the surroundings. Some types of outer jackets are made of shiny material to minimise heat transfer by radiation, since shiny materials are poor absorbers and radiators of heat (chapter 14).

CHAPTER 14

Practical activity 13.1

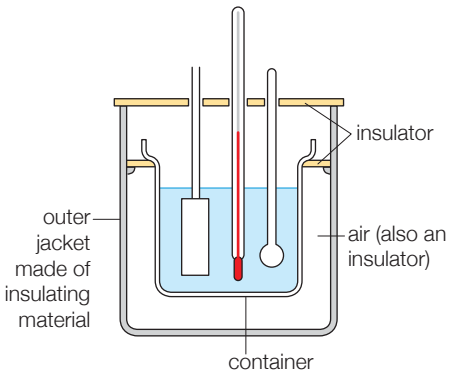


Figure 13.9 The container is thermally insulated.

Determining the specific heat capacity of a liquid by an electrical method



Caution: Electrical hazard



Caution: Heating hazard

The topic of electricity is treated in section D. However, this activity is simple enough for you to put together at this stage, with a little assistance from your teacher.

Since heating, electricity and liquid are involved, you must consult with your teacher concerning precautions to be taken concerning the method of setting up the apparatus, connecting the circuit and using the equipment.

You are to design and carry out an experiment to determine the specific heat capacity of a given liquid by an electrical method and to state whether the liquid given was most likely water. You are to use a metal calorimeter.

The designing is best done as a group activity. You are advised to study carefully the preceding section 'Measuring the specific heat capacity of a liquid by an electrical method' and Worked example 13.2.

The design

1 Examine and list the equipment to be used.

2 Draw and label the experimental arrangement, including the circuit diagram.

3 Outline the method in steps. (Point out precautions to be taken for safety at various steps and to obtain accurate results.)

4 Prepare a format for collecting data.

5 State the equation(s) you plan to use.

6 Discuss your written plan with your teacher.

The activity and the report

1 Carry out the activity under the watchful supervision of your teacher.

2 Write up your lab report. (Make sure you pay attention to units and significant figures in your data and calculations, and point out sources of possible experimental error.)

3 Discuss whether the liquid given was most likely water.

Extension

Find out how the specific heat capacity of water:

(a) helps living organisms to survive variations in temperature;

(b) helps to prevent motor car engines from overheating;

(c) helps islands, such as The Northern Bahamas, to have a pleasant climate during winter months.

Worked example 13.2: Specific heat capacity of a liquid by the direct (electrical) heating method

The following data were obtained in a direct electrical heating experiment (see figure 13.7) to find the specific heat capacity, c, for water. Determine the value of c from the data.

Electric current, I	= 1.5 A
Voltage across heater, V	= 12 V
Time current switched on, t	= 20 min 00 s = 1200 s
Mass of polystyrene container, m_1	= 5.0 g = 0.005 kg
Mass of polystyrene container + water	= 260 g
Therefore, mass of water, m	= 255 g = 0.255 kg
Initial temperature of water (and container), θ_1	= 15°C
Final temperature of water (and container), θ_2	= 35°C

Heat supplied by heater = heat gained by water and container

Therefore

$$IVt = mc(\theta_2 - \theta_1) + m_1c_1(\theta_2 - \theta_1)$$
$$\approx mc(\theta_2 - \theta_1)$$

As m_1 is small as compared with m, the term m_1c_1 is quite small compared with mc. In other words, the heat capacity of the polystyrene container is small compared to that of the water. Thus, when polystyrene containers are used in such experiments, their thermal contributions are usually ignored in the calculations, as we have done here. So we have

$$1.5 \times 12 \times 1200 = 0.255 \times c \times (35 - 15)$$

$$21\,600 = 5.1c$$

$$c = 4200 \text{ J K}^{-1}\text{kg}^{-1} \text{ (to 2 sig. fig.)}$$

ITQ4

What physical property of polystyrene leads us to assume that the mass of a polystyrene container would be small?

ITQ5

Give two reasons why the heat capacity of a copper calorimeter is large (and therefore should not be ignored in calculations).

Measuring the specific heat capacity of a solid metal by an electrical method

The specific heat capacity of a solid metal can be found experimentally by a similar direct heating method. (See figure 13.10 for the experimental arrangement.) The heat supplied by the heating coil is equal to the heat gained by the metal block. For a block of mass m and specific heat capacity c,

$$IVt = mc(\theta_2 - \theta_1)$$

The value of c is calculated using the known values of the other quantities in the equation.

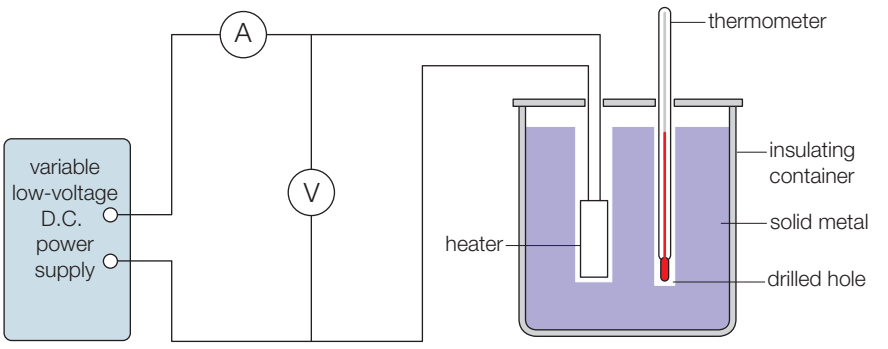


Figure 13.10 Determining the specific heat capacity of a metal by an electrical method.

Practical activity 13.2



Caution: Heating



Caution: Electrical hazard

Determining the specific heat capacity of a metal solid by an electrical method

The topic of electricity is treated in section D. However, this activity is simple enough for you to put together at this stage, with a little assistance from your teacher. Since heating and electricity and liquid are involved, you must consult with your teacher, as in Practical activity 13.1, concerning the precautions to be taken concerning the method of setting up the apparatus, connecting the circuit and using the equipment.

You are to design and carry out an experiment to determine the specific heat capacity of a given metal solid in the form of a block (figure 13.10) by an electrical method and, using this result, to suggest the material of which the metal is most likely made.

The designing is best done as a group activity. You are advised to study carefully the preceding section 'Measuring the specific heat capacity of a solid by an electrical method'. You may use a reporting format similar to that of Practical activity 13.1.

Questions

1 Why does the metal block in this experiment need an insulating container?

2 Would it be better for the heating to be done for a longer or shorter duration? Explain your answer.

Finding the specific heat capacity of a solid or liquid by the method of mixtures

method of mixtures ›

In the '**method of mixtures**' (figure 13.11), a metal solid, heated in boiling water, is quickly transferred to a liquid in a Styrofoam (polystyrene) cup. The mixture is stirred gently until a steady temperature is reached. The specific heat capacity of the metal (or the liquid) can be found using the equation:

heat lost by metal = heat gained by liquid

A Rumford correction could be applied in the experiment by starting with the liquid a few degrees below room temperature in an attempt to end with the mixture at approximately the same number of degrees above room temperature. However, since metals are very good conductors of heat, the metal (as well as the liquid) quickly reaches the equilibrium temperature. Heat loss to the environment is small since the process is of such short duration. Hence, the correction would not make much difference to the final experimental result.

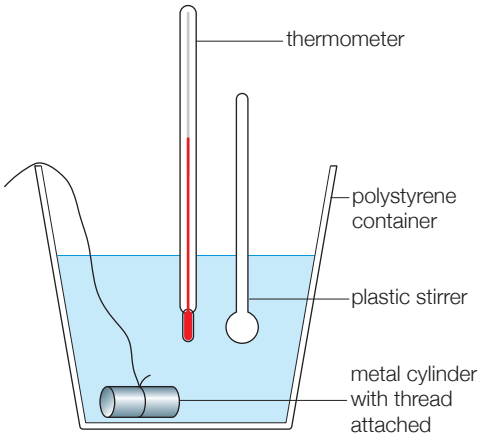


Figure 13.11 Determining specific heat capacity by the method of mixtures. (See SBA exercise 13.1.)

Worked example 13.3: Specific heat capacity of a liquid by the method of mixtures

The following results were obtained in an experiment to find the specific heat capacity of a liquid. Determine c_1 from the data.

Mass of polystyrene container	= 10 g = 0.010 kg
Mass of container + liquid	= 465 g = 0.465 kg
Therefore, mass of liquid, m_1	= 455 g = 0.455 kg
Initial temperature of liquid, θ_1	= 25°C
Final temperature of liquid, θ_2	= 33°C
Specific heat capacity of liquid, c_1	= ?

Mass of metal, m_2 = 200 g = 0.200 kg
Initial temperature of metal, θ_3 = 100°C
(assumed equal to the temperature of boiling water in which it was heated)
Final temperature of metal, θ_2 = 33°C
Specific heat capacity of metal, c_2 = 900 J K^{-1} kg^{-1}

heat lost by metal = heat gained by liquid

(we have ignored the heat gained by the container, see ITQ6). Therefore

$$m_2 \times c_2 \times (\theta_3 - \theta_2) = m_1 \times c_1 \times (\theta_2 - \theta_1)$$
$$0.200 \times 900 \times (100 - 33) = 0.455 \times c_1 \times (33 - 25)$$

from which we get c_1 = 3313 J (C°)$^{-1}$ kg^{-1}
= 3300 J K^{-1} kg^{-1} (to 2 sig. fig.)

(Alternatively, if we did our calculations using mass in g, our answer would have been 3.3 J K^{-1} g^{-1}, which we could then simply multiply by 1000 to obtain the answer in the S.I. unit of J K^{-1} kg^{-1}.)

ITQ6
In this experiment why can we ignore the heat gained by the polystyrene container?

Practical activity 13.3

Determining the specific heat capacity of the metal solid in a coin using the method of mixtures

Design and carry out an investigation, involving the method of mixtures, to find out the metal most likely used in making a certain metal coin (or a metal bolt). You are advised to study the section 'Finding the specific heat capacity of a solid or liquid by the method of mixtures' and Worked example 13.3. Your planning and reporting may follow the patterns used in Practical activities 13.2 and 13.3.

Specific latent heat

Change of state

CHAPTER 11

In chapter 11 (pages 174–175), we learned that when a substance changes its state at its boiling point and at its melting point, the temperature remains constant during the change. Figure 13.12 shows how the temperature of a beaker of crushed ice changes with time when it is heated. The graph shows that the temperature remains constant while the ice is melting to become water and also while the water is vaporising into steam. The temperature remains constant even though more heat is continually being supplied. Why is this?

During the change of state from ice to water, heat energy is used to break the bonds that hold the water molecules in the crystalline ice structure. The molecules do not vibrate any faster, so the temperature does not rise; they only become more disordered until all the solid has been converted to liquid. Since the heat energy supplied does not lead to a change of temperature during the change of state, the heat energy is called 'hidden heat' or **latent heat**.

latent heat

Similarly, when the water is changing into steam, all the heat energy that is added is used to do work against the attractive forces that keep the molecules of liquid together. Again, the overall kinetic energy of the molecules does not increase during this process, so we do not see a rise in temperature during the change of state.

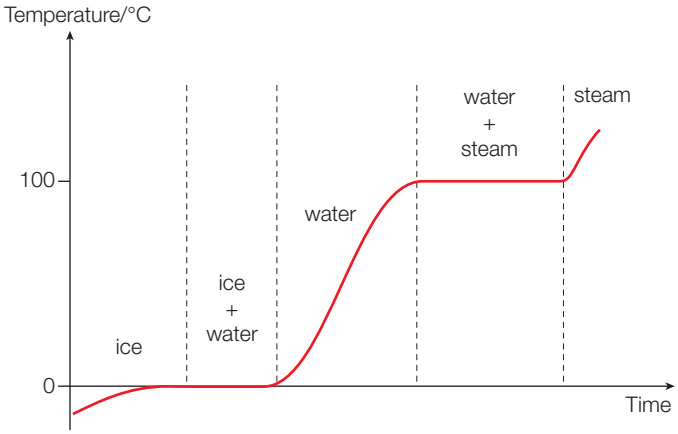


Figure 13.12 Temperature versus time graph for heating of crushed ice.

Specific latent heat of fusion

specific latent heat of fusion

The S.I. unit for specific latent heat of fusion is J kg^{-1}.

The specific latent heat of fusion, l_f, of a solid is the quantity of heat required to change 1 kg of the solid into liquid without a change of temperature:

$$l_f = \frac{E_H}{m} \text{ J kg}^{-1}$$

Here E_H is the quantity of heat that changes m kg of the solid into liquid without change of temperature. Rearranging, we get

$$E_H = ml_f$$

From the law of conservation of energy, l_f is also equal to the heat that must be removed from 1 kg of the liquid to change it to a solid without a change in temperature.

Figure 13.13 shows an apparatus for finding the specific latent heat of fusion of ice using the method of mixtures. Melting ice is added to a known mass of water at a known temperature. When all the ice has melted, the final temperature is recorded and the mass of the cup + water + melted ice is measured. The heat lost by the water originally in the cup can be calculated, and this is equal to the heat used to change the ice into water at 0°C to the final temperature.

Some of the precautions taken are as follows:

1 Pieces of melting ice rather than crushed ice are used. The temperature of the ice can then be assumed to be 0°C.

2 The ice is dried using tissue paper and quickly placed into the water. This is to ensure that ice, rather than ice and water, is being added to the water in the container.

3 To minimise net heat exchange with the atmosphere, water at a few degrees above room temperature is used in the container. Ice is added until the temperature of the mixture is the same number of degrees below room temperature (see Rumford's cooling correction, page 200).

The calculation method is illustrated in Worked example 13.4.

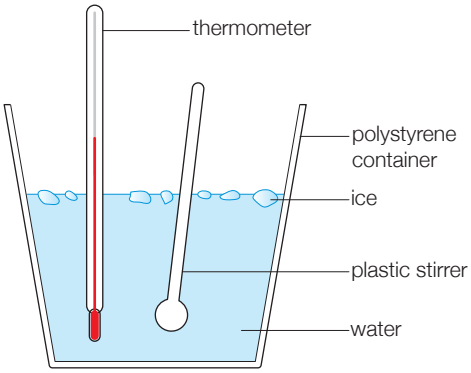


Figure 13.13 Experimental determination of the specific latent heat of fusion of ice using the method of mixtures.

Worked example 13.4: Specific latent heat of fusion of ice (method of mixtures)

The following results were obtained in an experiment to find the specific latent heat of fusion of ice, l_f, by the method of mixtures (see figure 13.13). Find l_f from the data.

Mass of polystyrene cup, m_1 = 5.0 g
Mass of polystyrene cup + water, m_2 = 335 g
Initial temperature of water, θ_1 = 33°C
Final temperature of water, θ_2 = 18°C
Specific heat capacity of water, c = 4.2 J k^{-1} g^{-1}
(as the masses involved are small, for convenience we are using joules per kelvin per gram)
Specific latent heat of fusion of ice, l_f = ? J g^{-1}
Initial temperature of crushed melting ice, θ_3 = 0.0°C
Final temperature of water formed from the melting ice, θ_2 = 18°C
Mass of cup + water + melted ice, m_3 = 390 g

We have

heat lost by water originally in container
= heat gained by melting ice
+ heat gained by water that resulted from melted ice

(due to its small heat capacity, we ignore the heat lost by the polystyrene container)

That is:

(mass of original water) × c × ($\theta_1 - \theta_2$)
= [(mass of ice) × l_f)]
+ [(mass of water formed from ice) × c
× (change from 0°C to final temperature)]

So we have

$$(m_2 - m_1) \times c \times (\theta_1 - \theta_2) = [(m_3 - m_2) \times l_f] + [(m_3 - m_2) \times c \times (\theta_2 - \theta_3)]$$

$$(335 - 5) \times 4.2 \times (33 - 18) = [(390 - 335) \times l_f] + [(390 - 335) \times 4.2 \times 18]$$

$$l_f = 300\,\text{J g}^{-1} \text{ (2 sig. fig.)}$$

ITQ7

In Worked example 13.4, what is the value of l_f, the specific latent heat of fusion of ice, in J kg^{-1}?

Practical activity 13.4

Specific latent heat of fusion of ice using an electrical method

Your group has access to the following laboratory equipment: variable low voltage D.C. supply, low voltage electric immersion heater, connecting wire, funnel, ice, beaker, balance and stop-clock. Design and carry out an investigation, involving an electrical method, to find the specific latent heat of fusion of ice. Your planning and reporting may follow the patterns used in Practical activities 13.2 and 13.3.

Specific latent heat of vaporisation

specific latent heat of vaporisation

The specific latent heat of vaporisation, l_v, of a liquid is the quantity of heat required to change 1 kg of the liquid to 1 kg of vapour without any change in temperature, at standard pressure:

$$l_v = \frac{E_H}{m}\ \text{J kg}^{-1}$$

Here E_H is the quantity of heat that changes m kg of the liquid into vapour without change of temperature. Rearranging the above equation, we get

$$E_H = ml_v$$

From the law of conservation of energy, l_v is also equal to the heat that must be removed from a vapour to change it to liquid without a change in temperature.

Figure 13.14 shows an experimental arrangement that can be used to determine the specific latent heat of vaporisation of water. A simple electrical method is shown.

An electric immersion heater is placed in the can of water. When the water is boiling, the balance is read and a stop-clock is started at the same time. After a time (t) has elapsed, the balance is read again. The difference in readings on the balance gives the mass (m) of water that was boiled off and converted to vapour during the time (t).

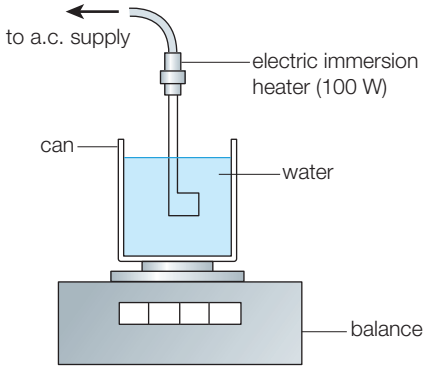


Figure 13.14 An electrical heating method for determining the specific latent heat of vaporisation of water.

Practical activity 13.5

Caution: Electrical hazard



Caution: Heating hazard



Caution: Wear safety goggles provided



Determining the specific latent heat of vaporisation of water

You will need:

- 100 W immersion heater (such as used for heating a cup of water)
- 400 ml beaker
- top-pan balance
- water
- stop-clock or stopwatch.

Method

1. Connect up the apparatus as shown in figure 13.14. The heater must be placed well into the water. (Caution: set up the apparatus under the direct supervision of your teacher. Both the electricity mains supply that supplies the heater and the heat generated are hazards.)
2. About 2 minutes after the water starts boiling, turn on the stop-clock and simultaneously note the reading on the balance.
3. When 20 g of water has boiled off, stop the stop-clock and note the time.
4. Using your measurements calculate the specific latent heat of vaporisation of water (see Worked example 13.5).

Questions

1. Why must the heater be placed well into the water?
2. State *two* sources of error in this experiment. Briefly describe how each error can be reduced.
3. Would a more accurate result be obtained if 40 g of water rather than 20 g were boiled off and used in the calculations? Discuss your answer.
4. Would the heat capacity of the container matter in this kind of experiment? Explain your answer.
5. Why are scalds from steam more serious than scalds from water boiling at the same temperature?
6. In what *two* ways can heat loss from the sides of the can be reduced?

Extension

Design an experiment to find the specific latent heat of steam, using a steam-generating can and a method of mixtures. State *two* precautions you would include in your design so as to obtain accurate results.

Worked example 13.5 shows how the value of l_v is calculated.

Worked example 13.5: Specific latent heat of vaporisation of water, l_v

The following data were obtained in an experiment to find the specific latent heat of vaporisation of water (see figure 13.14).

Mass of can + water at start, m_1 = 255.0 g
Mass of can + water at end, m_2 = 245.0 g
Power of heater, P = 100 W
Time heater was on, t = 3 min 50 s = 230 s

Heat energy supplied by heater = heat energy used in converting water into steam

CHAPTER 9

Since (see chapter 9, page 136)

$$\text{power} = \frac{\text{work done}}{\text{time taken}} = \frac{\text{energy}}{\text{time}}$$

$$\text{energy supplied} = \text{power} \times \text{time}$$

So,

$$\text{power} \times \text{time} = \text{mass of water converted to steam} \times l_v$$

$$P \times t = (m_2 - m_1) l_v$$

$$100 \times 230 = (255.0 - 245.0) l_v$$

$$23\,000 = 10.0 l_v$$

$$l_v = 2300\ \text{J g}^{-1} = 2\,300\,000\ \text{J kg}^{-1}$$

Refrigeration and air-conditioning

volatile

Refrigerators and air-conditioners make use of the latent heat of vaporisation of liquids. A **volatile** liquid (one that evaporates easily) is circulated in a closed system of metal pipes.

The volatile liquid used in a refrigerator is called a refrigerant. The use of refrigerants made of chlorofluorocarbons (CFCs) has been banned internationally, because they damage the Earth's protective ozone layer when they escape into the atmosphere.

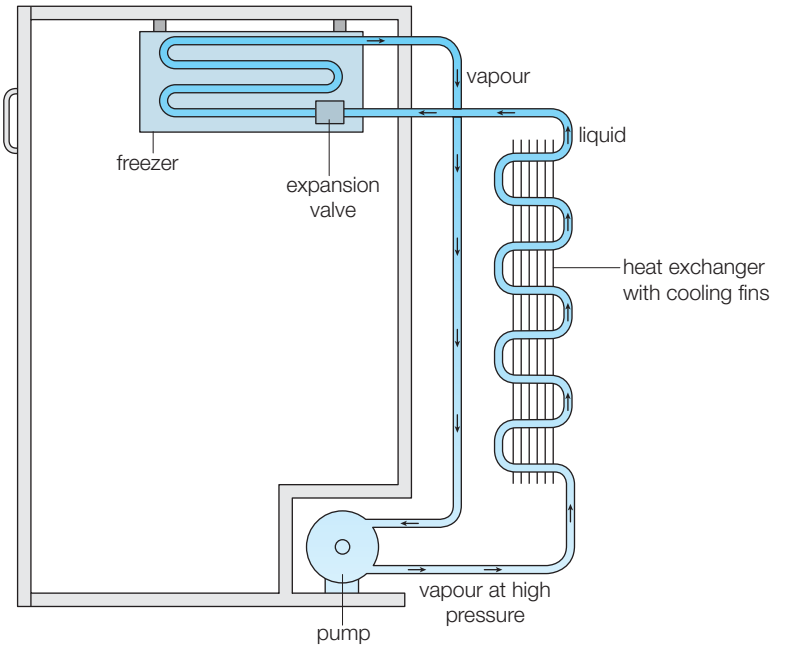


Figure 13.15 How a refrigerator works.

expansion valve ›

In the refrigerator (figure 13.15), the volatile liquid is forced through a tiny hole called an **expansion valve**, which is located in the system of pipes in the freezer section of the refrigerator. Expanding through the valve, the liquid turns to vapour. The conversion from liquid to vapour needs latent heat. This heat energy is taken from the system of pipes in the freezer. The pipes system in turn takes heat from the food and air present in the freezer section, by conduction and convection. In this way, the food within the freezer becomes cold. The main refrigerator section becomes cooler mainly by convection (see chapter 14).

ITQ8
Why does an air-conditioning unit function more efficiently when placed in a shaded area?

CHAPTER 14

The vapour leaving the freezer section by means of the system of pipes is compressed by the pump and becomes liquid. During the conversion from vapour to liquid, the vapour gives up latent heat to the condenser pipes and metal cooling fins, causing them to become hot. These in turn lose heat to the atmosphere mainly by radiation and convection (see chapter 14).

For air-conditioning, the cooling fins are placed outside the room (see figure 13.16). The latent heat in this case is taken from within the room, and so the room gets cool. For most efficient functioning, the air-conditioning unit is placed in a shaded part of the building.

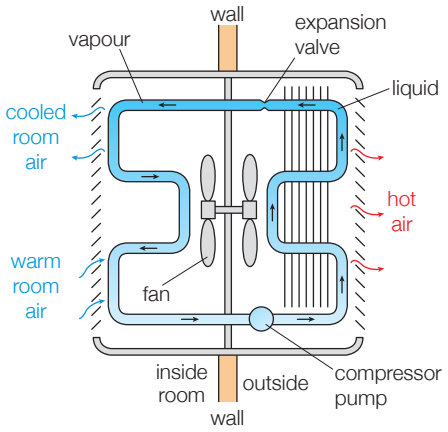


Figure 13.16 An air-conditioning unit.

ITQ9
Most of our body is water. Approximately how many joules of energy will therefore be required to raise the temperature of a 50 kg student by 1 K?

The importance of specific heat capacity and latent heat

Look at table 13.2 and compare the values for the specific heat capacity and the specific latent heat of water. The specific heat capacity of water is large: it takes 4200 J of energy to raise the temperature of 1 kg of water by just 1 K.

ITQ10
Give another reason for making cooking pots from metal, apart from low specific heat capacity.

Table 13.2 Thermal data for water (to 2 sig. fig.).

Specific heat capacity of water	4 200 $J K^{-1} kg^{-1}$ = 4.2 $J K^{-1} g^{-1}$
Specific latent heat of fusion of ice	330 000 $J kg^{-1}$ = 330 $J g^{-1}$
Specific latent heat of vaporisation of water at standard pressure	2 300 000 $J kg^{-1}$ = 2 300 $J g^{-1}$

The large specific heat capacity of water helps to keep our body temperatures from rising too much when the temperature of the surroundings increases. Water is also used to fight most fires because it absorbs a lot of heat before it undergoes a significant rise in temperature. The water lowers the temperature of the **combustible** materials to below their **ignition temperature**. Water is also used as a coolant in the radiators of car engines, since it can absorb a lot of heat without undergoing a large rise in temperature. This helps to keep the engine from overheating.

combustible ignition temperature ›

ITQ11
Which is likely to take a longer time, cooking an egg in an iron pot or in an aluminium pot of the same mass, starting from room temperature? Explain your answer. (Hint: see table 13.1.)

Metals like copper, iron and aluminium, on the other hand, have relatively low specific heat capacities (390, 450 and 900 $J K^{-1} kg^{-1}$, respectively). Cooking pots are made of metal partly because a relatively small quantity of heat results in a large temperature rise in the pot, so heat is not wasted just to heat up the pot to the desired cooking temperature. (Of course, they are also used because of their high melting points and good thermal conduction properties.)



Figure 13.17 It takes only two ice cubes to cool down a soft drink noticeably.

BIOLOGY: homeostasis

The specific latent heat of fusion of ice is about 80 times that of the specific heat capacity of water. Thus, two ice cubes can cool a glass of water considerably, since it takes 333 J of energy to melt 1 g of ice, but only 4.2 J of energy to change the temperature of 1 g of water by 1 K (figure 13.17). The energy to melt the ice comes from the water into which the cubes are put. As the water loses heat energy, its temperature drops.

The specific latent heat of vaporisation of water is 2 260 000 J kg^{-1}, that is, about 500 times the specific heat capacity of water. Thus, when steam condenses to form water, it gives out a considerable quantity of heat. For this reason, steam scalds are much more serious than scalds caused by boiling water at the same temperature. When sweat evaporates, it takes a large amount of heat from our skin since the specific latent heat of vaporisation of water is so high. The *evaporation of sweat* therefore produces a cooling effect and helps to prevent our bodies from overheating.

Chapter summary

- According to the Caloric Theory, heat was thought of as a fluid substance. A gain (or loss) of caloric caused the temperature of an object to rise (or fall).
- The Kinetic Theory of heat regards heat as a net transfer of average kinetic energy of molecules. The direction of the net transfer is from an object at a higher temperature to one at a lower temperature. Experiments conducted by Count Rumford and James Joule helped to refute the Caloric Theory and confirm the energy transfer theory of heat.
- The specific heat capacity, c, of a substance is defined as the quantity of heat, E_H, that raises the temperature of 1 kg of the substance by 1 K.
  $c = \frac{E_H}{m\Delta T}$ J K^{-1} kg^{-1}
- The heat capacity, C, of an object is defined as the quantity of heat required to raise the temperature of the object by 1 K.
  $C = \frac{E_H}{\Delta T}$ J K^{-1}
- The specific latent heat of fusion, l_f, of a solid is defined as the quantity of heat required to change 1 kg of the solid into liquid without a change of temperature.
  $l_f = \frac{E_H}{m}$ J kg^{-1}
- The specific latent heat of vaporisation, l_v, of a liquid is the quantity of heat required to change 1 kg of the liquid to 1 kg of vapour without any change in temperature, at standard pressure.
  $l_v = \frac{E_H}{m}$ J kg^{-1}
- Specific heat capacities and specific latent heats can be determined by electrical methods or by methods of mixtures.

Answers to ITQs

ITQ1 $E_H = mc\Delta\theta = 5.0 \times 4200 \times 10 = 210\,000$ J

ITQ2 Water.

ITQ3 $C = mc = 0.200 \times 900 = 180$ J K^{-1}

ITQ4 Polystyrene has a very low density.

ITQ5 A copper container has a relatively large mass and copper has a significant specific heat capacity. Thus, $C = mc$ is large and should not be ignored.

ITQ6 Both m and c are relatively small for the polystyrene container. Hence $C = mc$ is small and can be ignored.

ITQ7 300 000 J kg^{-1} (since 1 J g^{-1} = 1000 J kg^{-1})

ITQ8 The temperature in the shade is lower than the temperature in the sunshine. Hence heat transfer will be greater from the air-conditioner to the shaded area than to an area in direct sunshine, resulting in better cooling by the air-conditioner.

ITQ9 $E_H = mc\Delta T = 50 \times 4200 \times 1 = 210\,000\,J$

ITQ10 Metals are good conductors of heat.

ITQ11 In the aluminium pot. At the start of the heating, the aluminium pot will need twice as much heat energy as the iron pot in order to achieve the same rise in temperature (since aluminium has about twice the specific heat capacity as iron).

Examination-style questions

1 100 megajoules is equivalent to

A 10^{-8} J B 10^{-6} J C 10^{6} J D 10^{8} J

Items 2 and 3

The following readings were obtained in an experiment to find the specific latent heat of vaporisation of a certain liquid:

Power of heater, $P = 100$ W

Mass of liquid boiled off, $m = 60$ g

Time taken, $t = 3$ min

2 The energy supplied by the heater is:

A 100 J B 300 J C 6000 J D 18 000 J

3 The specific latent heat of the liquid could be expressed in:

A W g^{-1} B W m^{-1} C J g^{-1} D J m^{-1}

4 Which scientist showed that a certain quantity of heat was equivalent to a fixed quantity of mechanical energy?

A Einstein B Joule C Newton D Rumford

Items 5 and 6

A 2 kg block of aluminium has specific heat capacity 900 J K^{-1} kg^{-1}. The temperature of the bock rises from 5°C to 10°C.

5 The heat capacity of the block is:

A 450 J K^{-1} B 1800 J K^{-1} C 2400 J K^{-1} D 9000 J K^{-1}

6 The quantity of heat supplied to the block is:

A 1800 J B 45 000 J C 9000 J D 18 000 J

7 Compared with a cup of water at room temperature, a red-hot spark has:

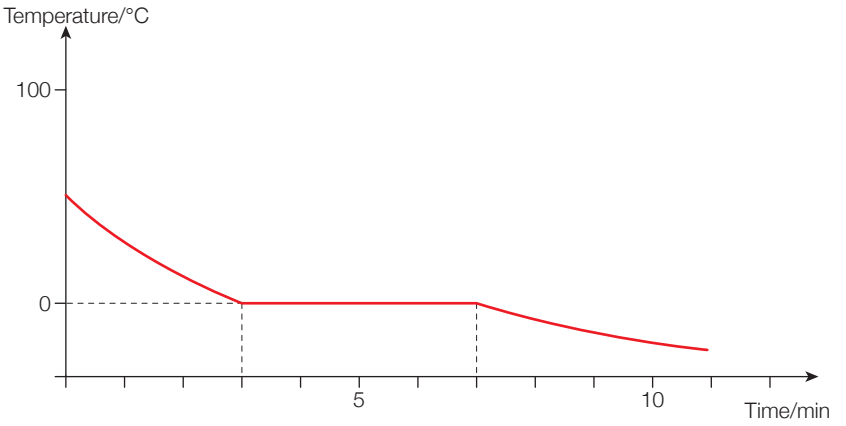
	Temperature	Thermal energy
A	lower	less
B	lower	greater
C	higher	less
D	higher	greater

8 Which of the following substances in contact with the skin is likely to give the most serious scalds (the substances have equal mass)?

A water at 100°C
B steam at 100°C
C iron at 100°C
D aluminium at 100°C

9 A container of water is cooled at a constant rate until all the water turns to ice. The graph below shows how the temperature changes with time.How many minutes did it take for all the water to become ice, starting from the time the conversion began?

A 3 B 4 C 7 D 8



10 When trains collide at high speeds, a fire is very likely the result. Explain how the fire can start, using the energy theory of heat.

11 (i) Draw and label a diagram of an electrical circuit that can be used to determine the specific heat capacity of a liquid.

(ii) Explain how the current can be kept steady in an experiment using the circuit in (i).

(iii) It is desired that 250 g of a water of specific heat capacity 4200 J K^{-1} kg^{-1} undergo a temperature rise of 5.0 K in 10 minutes.

(a) How much thermal energy is required for the temperature rise of 5.0 K?

(b) Determine the approximate power of an electrical heater that can be used in (a).

(c) In practice, the water takes a time longer than 10 minutes, using a heater with a calculated power as in (b). Suggest a reason for this.

12 (i) A metal bar becomes hot when left out in the sunshine. How could the rise in temperature be explained using:

(a) the Caloric Theory of heat?

(b) the modern Kinetic Theory of heat?

(ii) How did the evidence from Joule's experiments help to convince scientists that heat was a form of energy?

13 (i) Explain the reason for the following steps in an experiment to find the specific latent heat of fusion of ice by the method of mixtures:

(a) A polystyrene cup is used.

(b) Individual pieces of ice, rather than crushed ice, are used.

(c) Melting ice is used.

(d) The ice is dried before placing it into the water.

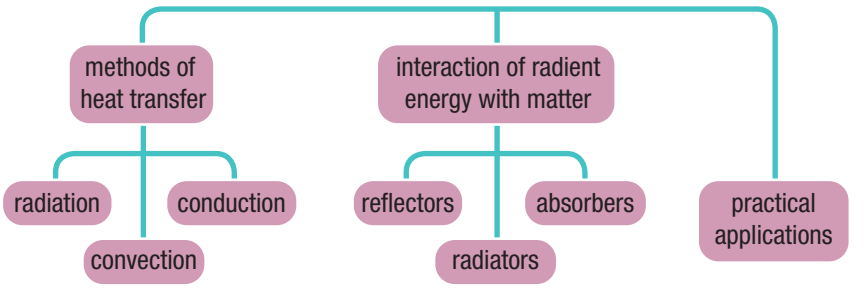
(e) Water at a starting temperature of a few degrees above room temperature is used.

(ii) In one such experiment, the initial and final temperatures of the water were 30°C and 20°C, respectively. If 25 g of ice gave rise to this change, calculate the total heat energy used in changing the ice to water at 20°C. (Use data as needed from tables 13.1 and 13.2, pages 199 and 209.)

14 Methods of Heat Transfer

By the end of this chapter you should be able to:

- explain the transfer of thermal energy (heat) by conduction
- describe an experiment to compare qualitatively the thermal conductivities of different solids
- describe an experiment to show that water is a poor conductor of heat
- relate the fact that air is a very poor conductor of heat to the insulation properties of certain materials
- explain the transfer of thermal energy by convection
- describe experiments that demonstrate convection in fluids
- relate convection to common phenomena
- recall that thermal energy can be transferred by electromagnetic (infrared) radiation
- describe an experiment to demonstrate that radiant energy (electromagnetic radiation) does not need a medium for its transmission
- describe experiments to investigate the factors on which the absorption and emission of radiation depend
- recall that good absorbers are also good emitters
- relate the phenomenon of radiation to everyday applications and experience
- explain the glasshouse (greenhouse) effect
- relate the principles of thermal energy transfer to the design of devices and homes



More ways than one!

CHAPTER 13

Heat energy is also referred to as thermal energy. In chapter 13, we discussed a theory that heat represents a net flow of vibrational energy of particles present within substances. The net flow of heat energy takes place from a hotter to a colder object until equilibrium is reached and both are at the same temperature. In figure 13.6, we illustrated this net transfer of vibrational energy involving two solids. This method of thermal energy transfer is called **conduction** and is characteristic mainly of solids.

conduction

In this chapter, we will also look at two other methods of heat transfer. In **convection**, the particles of a hot substance move physically from one region to another, carrying vibrational or translational energy with them.

convection

This method of thermal energy transfer is characteristic of fluids (liquids and gases). Energy may also be transferred by a method that is independent of matter, that is, by electromagnetic waves. This type of thermal energy transfer is called **radiation** and the energy is called **radiant energy**.

radiation ›
radiant energy ›

These three methods of heat transfer have many consequences and many useful applications (examples are shown in figures 14.1 and 14.2).



Figure 14.1 This type of environmentally friendly solar water-heater, found throughout the Caribbean, uses three methods of heat transfer.



Figure 14.2 Both the rug and the wooden floor are at the same temperature in the early morning. Yet, the wooden floor feels much colder than the rug. Why is this? You will find the answer in this chapter.

Methods of heat transfer

Heat energy represents a net transfer of energy from a hotter to a colder object. The transfer may take place by conduction, by convection or by radiation.

Conduction

In conduction, the heat energy travels *along* matter. Figure 14.3 shows one end of a metal bar in contact with a flame. A thumbtack is stuck near the other end of the bar with some wax. After a while, the wax melts and the thumbtack falls. This shows that the other end of the bar has become hot. Heat energy has travelled along the bar from the flame towards the other end by conduction.

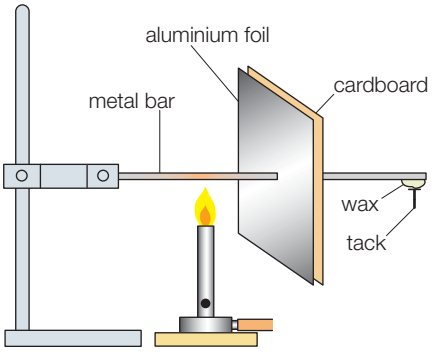


Figure 14.3 Demonstrating heat travel by conduction along a metal bar.

ITQ1

What is the purpose of the cardboard and foil in figure 14.3?

To explain conduction in a solid, we consider the particles making up the solid. In a non-metallic solid, these are usually molecules. The particles are constantly vibrating about a fixed position relative to each other. The molecules near the source of the heat begin to vibrate more rapidly. When they do so, they also cause the neighbouring molecules to vibrate more rapidly. The faster vibrations are passed on from molecule to molecule through the solid, to the cooler parts; this is quite a slow process. Thus, in conduction, the heat energy is passed from molecule to molecule, but there is no flow of the molecules themselves.

free electrons ›
ions ›

Metals are very good conductors of heat, however, because they contain many '**free electrons**' that can move through the fixed lattice structure of the vibrating metal **ions** (figure 14.4). As the ions near the heated end of the metal rod in figure 14.3 vibrate more quickly, they 'bump' into free electrons, which therefore gain translational kinetic energy. As these free electrons migrate in the lattice, they collide with ions along the way, causing the ions to vibrate faster. Hence the heat energy is transferred from the heated end of the rod quite quickly.

Heat conduction can also take place in liquids and gases: heat energy can be passed from molecule to molecule, as in the case of solids. However, liquids tend to be poorer conductors of heat than solids, though some liquids do conduct better than some solids. The average distance between molecules is greater in a liquid, and also they are not in fixed positions relative to each other; they are moving about. Thus energy is passed on less quickly from

molecule to molecule through the liquid than through the solid. The molecules of a gas are far apart from one another, and are also travelling in random directions. Transfer of heat energy from one molecule to another can only happen when there is a chance collision. Gases therefore tend to conduct heat *very poorly* compared with liquids and solids.

An ion is an atom that has gained or lost electrons. In metals, atoms lose their outer electrons easily forming a lattice of ions. The electrons move about this lattice very easily.

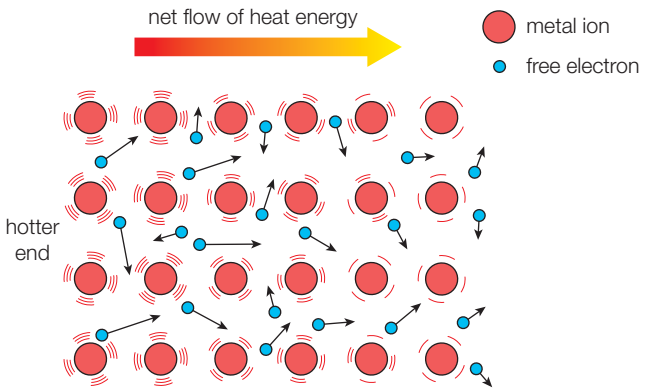


Figure 14.4 Electrons colliding with the ionic lattice structure of a solid metal.

It is no accident that metals that are good conductors of heat are also good electrical conductors. Both processes depend on the action of free electrons. Since poor electrical conductors do not have many free electrons, they cannot be good heat conductors.

Conductors

conductor ›

Metals are very good **conductors** of heat, because of their free electrons, as mentioned above. We say that they have a high **thermal conductivity**.

thermal conductivity ›

Figure 14.5 shows an experiment that can be done to compare the thermal conductivities of solids. The rods are of same length and diameter. Matchsticks are stuck on the rods, using equal quantities of Vaseline, at equal distances from the heat source. The matchstick that falls first indicates which material has the highest thermal conductivity.

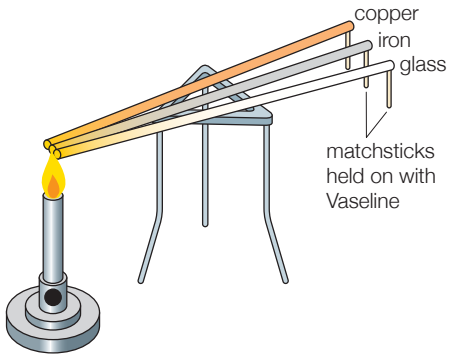


Figure 14.5 Comparing the thermal conductivities of solids.

Practical activity 14.1



Caution: Hot surfaces



Caution: Flammable gas

Comparing thermal conductivities of selected metals

1 Set up an apparatus as shown in figure 14.5. (Make sure there are no combustible materials such as loose paper, clothing or hair that can come into contact with the Bunsen flame.) Draw a diagram of the arrangement.

2 Measure and record the time taken for each matchstick to fall.

3 Repeat step 2 five times and determine and record average times.

Questions

1 Explain why 'the matchstick that falls first indicates which material has the highest thermal conductivity'?

2 Rank the thermal conductivities of the materials starting with the highest. Compare the ranking you obtained with thermal conductivity data of each material.

3 The flame itself gives off heat radiation which can reach the matches. Draw a diagram showing how the apparatus can be modified to minimise this possible source of error.

Insulators

insulator ›

Poor conductors of heat are called **insulators**. Thus, most non-metals, such as plastic and wood, are called insulators. Insulators can be solid, liquid or gas.

The experiment illustrated in figure 14.6 shows that water is a very poor conductor of heat. Although the heat causes the water near the top of the test tube to boil for several minutes, the ice at the bottom does not melt for quite some time, showing that hardly any heat is reaching the ice by conduction.

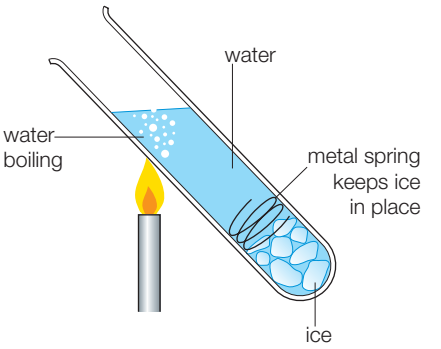


Figure 14.6 Showing that water is a poor conductor of heat.

Figure 14.7 illustrates another experiment that shows that water is a poor conductor of heat. An electric heater is immersed in the water near the top of a 250 ml measuring cylinder and the apparatus is kept still. With the heater on for about 5 minutes, the temperature is taken near the bottom and then quickly near the top. The temperature at the top is much higher than that at the bottom, showing that water does not conduct heat very well.

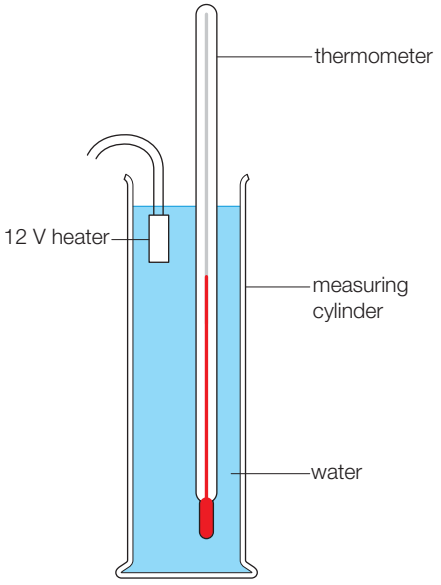


Figure 14.7 Investigating heat conduction in water.

IT Q2

Why does the wooden floor in figure 14.2 feel colder than the rug although they are both at the same temperature?

Air is a poor conductor of heat. Styrofoam (polystyrene) cups have many pockets of air in their walls. The air trapped in the walls is such an effective insulator that one can quite comfortably hold with bare hands a Styrofoam cup containing hot coffee. Birds fluff their feathers to trap air when they wish to keep warm. In cold countries, windows are made with two panes of glass enclosing a layer of air between the panes. This is called double-glazing. The air serves as an insulator and helps to reduce the flow of heat energy by conduction from the inside of a house to the outside. Air present in hollow blocks in the walls of these houses serves a similar purpose.



Figure 14.8 An Inuit snow house (called an igloo) provides adequate protection during severe cold.

IT Q3

Why is snow such a useful material thermally for building houses in Arctic regions (figure 14.8)?

Convection

convection ›

The experiment shown in figure 14.9 shows that heat energy can also be carried along with particles of matter. This method of heat transfer is called **convection**.

When the air near the candle gets hot, it expands, because the molecules are moving more quickly and push one another further apart. The air near the candle is now less dense than the cooler air around it and rises, the particles carrying their heat energy with them. The cooler air sinks and flows in to replace the hot rising air. This flow is demonstrated by the smoke in figure 14.9, and is called a **convection current**.

convection current ›

Supermarkets make use of the fact that colder air is more dense than warmer air when top-opening meat freezers are left open. The denser cold air remains in contact with the food and little heat is lost by convection.

Convection also takes place in liquids. Figure 14.10 shows convection in water. The hotter rising water carries heat with it. The flow of the water is shown by the colour of the potassium permanganate (potassium manganate (VII)).

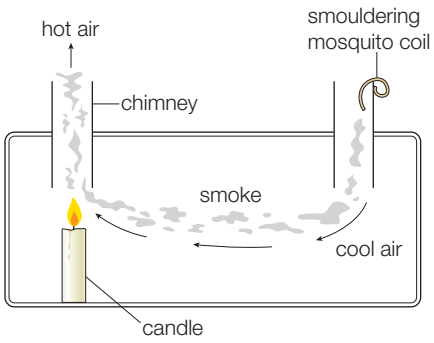


Figure 14.9 Demonstrating convection in air.

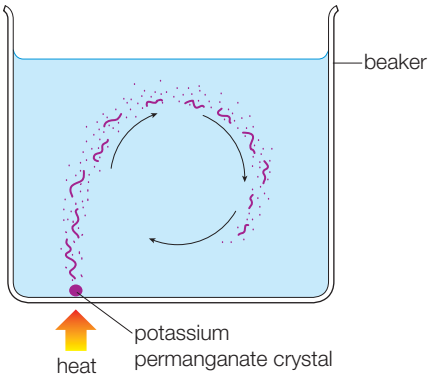


Figure 14.10 Demonstrating convection in water. The arrows show the direction of the convection currents.

ITQ4

At 10 a.m. one day, there was very little breeze on an island. Suggest a reason for this.

Land and sea breezes are convection currents of air. During the day, the land heats up faster than the sea, which does not change its temperature much. Hot air from the land rises and cooler air from the sea rushes in to take its place, forming a sea breeze. In the evening, the land cools faster than the sea. Warmer air begins to rise from above the sea, producing a breeze from the land.

ITQ5

Heat transfer by convection takes place in gases and liquids. Can convection occur in solids? Give a reason for your answer.

Domestic hot-water systems make use of convection, too. In one design the electrical heating element is placed near the bottom of the tank of water (figure 14.11). The water heated by the coil rises and carries heat to the top by convection. (Note that the outside of the tank is lined with fibreglass, which has pockets of trapped air. This lining serves as an insulator to reduce heat loss from the hot container, which would have been in direct contact with the moving air of the room.)

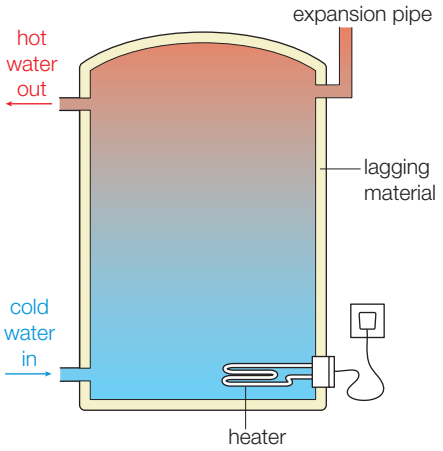


Figure 14.11 In this design, an immersion heater is positioned near the bottom of the domestic hot-water tank.

Radiation

If you place your hand near to the bottom of a cup containing a hot drink, you can feel the heat given off by the cup (figure 14.12). The heat is not likely to have travelled from the cup to the hand by conduction since the air present between the two is a poor conductor. Neither could the heat have reached the hand by convection, since hot air rises. The heat travelled by a process called **radiation**, involving electromagnetic (infrared) waves. Similarly, heat from a campfire reaches campers seated on the ground by radiation.

radiation ›

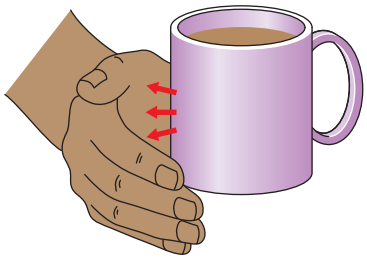


Figure 14.12 Heat energy travels by radiation from the cup to the hand.

IT Q6

If all objects above absolute zero of temperature are constantly giving off electromagnetic energy, then how can an object remain at a constant temperature in a room?

Radiant heat energy is a form of electromagnetic (e.m.) radiation, like light, radio waves or X-rays. (Electromagnetic radiation is discussed more fully in chapter 19.) Unlike conduction or convection, matter is not needed to transfer radiant energy. Heat energy from the Sun reaches the Earth mainly by radiation, since there is mostly empty space between the Sun and the Earth.

It has been shown that all objects above absolute zero of temperature emit electromagnetic radiant heat energy. The hotter the body, the more energy it radiates in a given time. There is a net flow of radiant heat energy from a hotter to a colder object.

CHAPTER 19

Our skin is very sensitive to infrared (IR) radiation (see electromagnetic spectrum, chapter 19, page 287). When this radiation falls on us, our skin feels hot and we describe this sensation as heat reaching our skin.

Figure 14.13 shows a photograph taken in total darkness by an infrared camera. The light areas are the parts of the scene that give off the most infrared radiation. The camera is able to 'see' in the dark by making use of the infrared electromagnetic waves given off by the objects in the scene. Some snakes can also follow the 'heat-print' tracks of their prey in total darkness because they can sense infrared radiation.



Figure 14.13 Photograph of a scene taken in total darkness using an infrared camera.

IT Q7

In figure 14.13, which is hotter, the wall or the edge of the roof?

Absorbers, reflectors and radiators of electromagnetic (radiant) energy

absorber | emitter | reflector

Matter may **absorb**, **reflect** or **emit** (**radiate**) electromagnetic radiant heat energy.

In the experiment shown in figure 14.14, the wax melts and the cork drops off first from the metal plate on the right. This suggests that black objects absorb radiant heat better than silvery or white objects. This could be explained by saying that silvery and white objects tend to reflect heat, whereas black objects tend to absorb heat. For this reason, in the sunshine, white clothing feels cooler than black clothing (figure 14.15).



Figure 14.15 Players wear white clothing during daytime cricket.

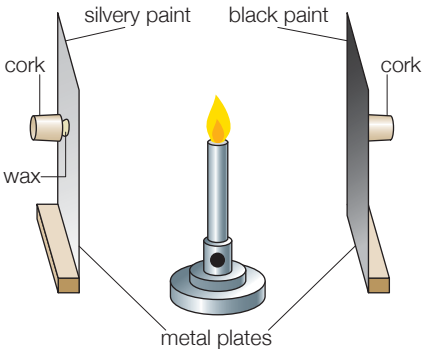


Figure 14.14 The metal plates are of the same size and material and are placed equal distances from a Bunsen flame.

Practical activity 14.2

Absorption of heat radiation



Caution: Hot surfaces



Caution: Flammable gas

(In this experiment, make sure there are no combustible materials such as loose paper, clothing or hair that can come into contact with the Bunsen flame.)

You may use an arrangement similar to that shown in figure 14.14.

You are to design experiments to investigate the following factors as they affect absorption of heat radiation:

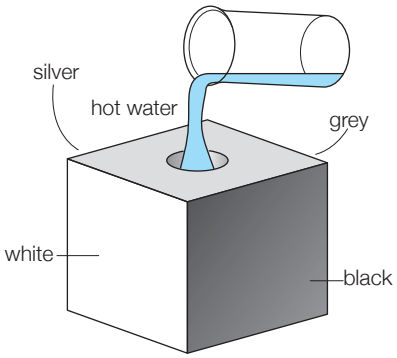


Figure 14.16 A Leslie's cube for investigating heat radiation by different types of surfaces.

(a) nature of surface – shiny or dull;
(b) texture of surface – rough or smooth;
(c) area of surface.

In your design, you must show clearly how you plan to control variables and exactly how you plan to achieve the various types of surfaces. You must also draw a diagram and show how you plan to record your data.

After you have discussed your written plan with your teacher, you should conduct the investigation and write up a laboratory report.

Figure 14.16 shows a metal cube containing hot water (called a Leslie's cube). The four faces are at the same temperature as the water. A thermometer, placed a suitable distance from each side will indicate, by its temperature rise, which surface will give off (i.e. radiate) heat better than another.

Practical activity 14.3

Emission of heat radiation



Caution: Hot surfaces



Caution: Flammable gas

(In this experiment, make sure there are no combustible materials such as loose paper, clothing or hair that can come into contact with the Bunsen flame.) You may use a Leslie's cube arrangement similar to that shown in figure 14.16. However, you may need to make modifications to some of the surfaces in order to investigate the factors listed below. You are to design experiments to investigate the following factors as they affect emission (i.e. radiation) absorption of radiant heat energy:

(a) nature of surface – shiny or dull;
(b) texture of surface – rough or smooth;
(c) area of surface;
(d) colour of surface – black or white.

In your design, you must show clearly how you plan to control variables and exactly how you plan to achieve the various types of surfaces. You must also draw a diagram and show how you plan to record your data.

Experiments have shown that black surfaces are good emitters of heat energy as well as being good absorbers. Radiators in motor cars and in refrigeration systems are painted black so that they can give off heat quickly and thus cool the systems faster.

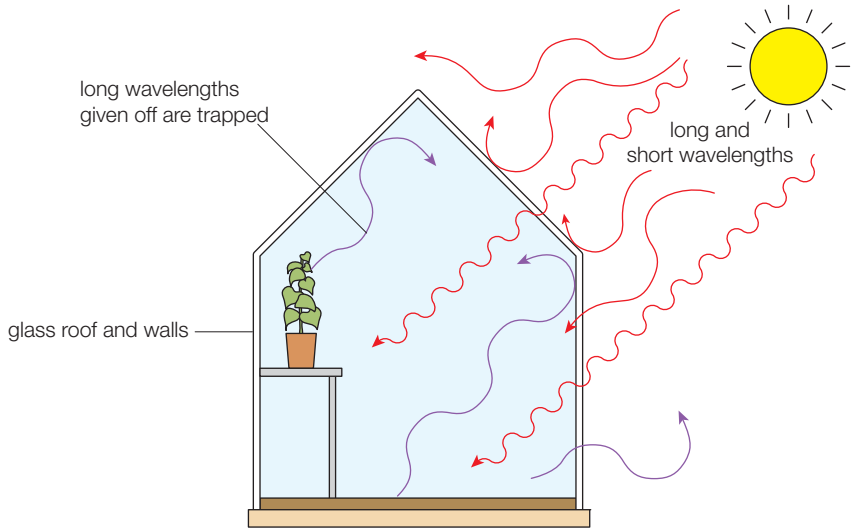


Figure 14.17 How a greenhouse traps heat radiation.

The 'greenhouse effect'

In winter, in cooler countries, some plants need to be at warmer temperatures in order to grow properly. Such temperatures can be obtained inside a greenhouse (sometimes called a glasshouse) even though it is cold outside. In a typical greenhouse, even though the temperature outside might be 15°C, the temperature inside can be as high as 25°C. This temperature is attained through absorption of electromagnetic radiation.

Figure 14.17 shows plants growing inside a greenhouse. Much of the greenhouse is made of glass. Short-wavelength infrared electromagnetic waves from the Sun pass easily through glass and their radiant heat energy is

absorbed by the objects inside the greenhouse. The objects become warm and therefore radiate electromagnetic waves in their turn, but emit mostly long-wavelength infrared waves. Glass is not very transparent to long-wavelength infrared radiation; a large percentage of these waves are reflected inside the greenhouse. Thus, heat energy is trapped within the greenhouse and the temperature inside rises. This is the '**greenhouse effect**'.

greenhouse effect

A similar effect is seen in cars left in sunshine with the windows closed. The temperatures inside become considerably higher than the outside temperature. (This can even happen in hazy weather.) Children left inside can suffer from heat exhaustion, which is why it is not safe to leave children unattended in cars.

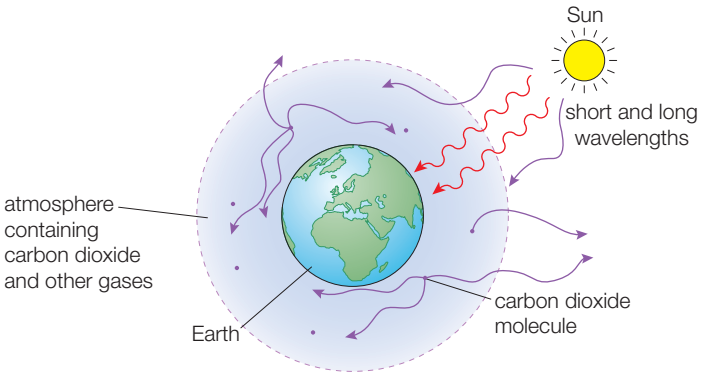


Figure 14.18 The 'greenhouse effect' gives rise to global warming.

The greenhouse effect has consequences for the Earth. Carbon dioxide (and other gases in the atmosphere) behave like the glass of the greenhouse, trapping heat energy from the Sun around the Earth. On account of its molecular structure, carbon dioxide absorbs and emits infrared radiation readily. Figure 14.18 shows electromagnetic radiation from the Sun, consisting of waves of long and short wavelengths, entering the Earth's atmosphere. Some of these are reflected back into space (not shown in diagram). Others heat up the Earth and atmosphere, giving rise to infrared radiation. Carbon dioxide readily absorbs this infrared radiation, but re-radiates it in all directions. Some is radiated out into space – the rest is trapped within the atmosphere or sent back to the Earth.

There is thus a net trapping of heat energy due to the presence of carbon dioxide in the atmosphere. This gives rise to the warming up of the whole Earth – a phenomenon we call **global warming**.

global warming

Too much carbon dioxide in the atmosphere, from burning fossil fuels such as oil and coal, causes the Earth to become warmer. This accelerates global warming. Global warming of even a few degrees Celsius causes the seas to expand and ice at the polar caps to melt, producing a rise in sea level and subsequent flooding in low-lying areas. Global warming is also thought to be responsible for some long-term changes in the climate in different regions of the Earth.

Practical applications of heat transfer methods

Cooking

Metal pots make use of the fact that metals are very good conductors of heat and non-metals, generally, are poor conductors. Thus, the body of a pot is made of metal, whereas the handle is usually insulated.

microwave

Microwave cooking makes use of electromagnetic radiation (figure 14.19). Microwaves (of wavelength approximately 12 cm) penetrate foodstuffs more deeply than the infra-red used in conventional cooking. As the wave penetrates, water in the

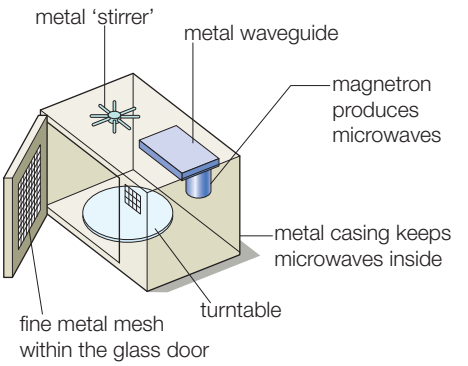


Figure 14.19 A microwave cooker.

food absorbs energy. The water molecules vibrate more vigorously and the temperature goes up. The food is more uniformly cooked, without the outer layer being crisped or browned by high temperatures.

Microwaves cannot pass through metal. The metal mesh behind the glass door of a microwave oven reflects the waves inside the oven. If the waves were to pass through the glass door, less energy would be available for cooking. The mesh also protects people from injury because microwave radiation damages living tissue. The rotating metal 'stirrer' reflects the microwaves produced by the magnetron so that they spread over the food evenly.

One of the reasons why microwave cooking is very efficient is that microwave energy is absorbed better by the food substances, which generally contain water molecules, than by the food containers, which do not. Thus, most of the energy goes to heating up the food rather than being wasted in heating up the food container.

Not all types of containers can be used safely in microwave cooking. Metals and certain types of ceramics are unsafe.

The vacuum flask

vacuum flask ›

The **vacuum flask** shown in figure 14.20 can keep its contents quite hot for more than a day.

The stopper is made of air-filled plastic or cork to reduce heat loss by conduction. The stopper also prevents hot air from escaping, so reducing heat loss by convection.

The double walls of the container are made of glass, itself a poor conductor, and there is a vacuum between them. Heat energy cannot travel by conduction across the vacuum, nor can heat be taken from the inner wall to the outer one by convection. The wall in contact with the contents is also silvered on both sides. This helps to reduce heat loss by reflecting back into the flask the electromagnetic waves that are given off by the hot contents. The outer wall of the container is also silvered on both sides. Thus, heat loss to the environment by radiation is reduced since shiny surfaces are poor radiators. What heat is emitted into the space between the walls bounces around this space by reflection off the silvered sides of the walls enclosing the vacuum. This further helps to minimise heat loss.

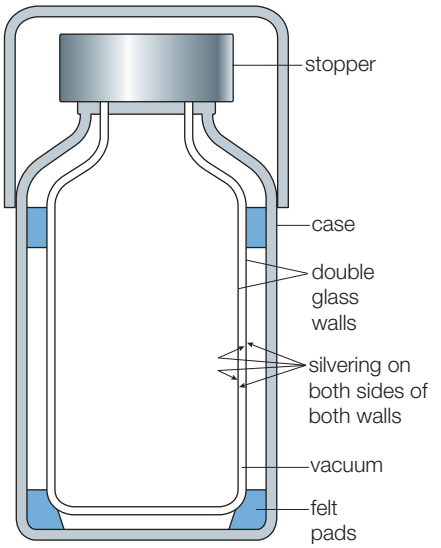


Figure 14.20 Structure of a vacuum flask.

The supports for the glass container are made of felt, a poor conductor of heat (The felt, being soft, also serves as a mechanical shock absorber for protection of the glass container.)

ITQ8

The vacuum flask can also keep a liquid cold. In what way does the silvering on the outer wall of the liquid container help to keep a liquid cold in a warm room?

The solar water-heater

solar panel ›

Figure 14.21 shows the essential features of a **solar panel** used as a water-heater. The *glass cover* traps heat in the same way as the glass in the greenhouse in figure 14.17. The *collector plate* is made of a very good conductor of heat such as copper or aluminium, and it is painted black so that it efficiently absorbs the solar radiation entering the panel. The plate either has water channels within it or a long, folded copper (or aluminium) tube welded on to it. The panel is backed by a sheet of *polyurethane* to reduce heat loss by conduction.

A photo of a different type of solar water-heater is shown in figure 14.1. (See also ITQ9.)

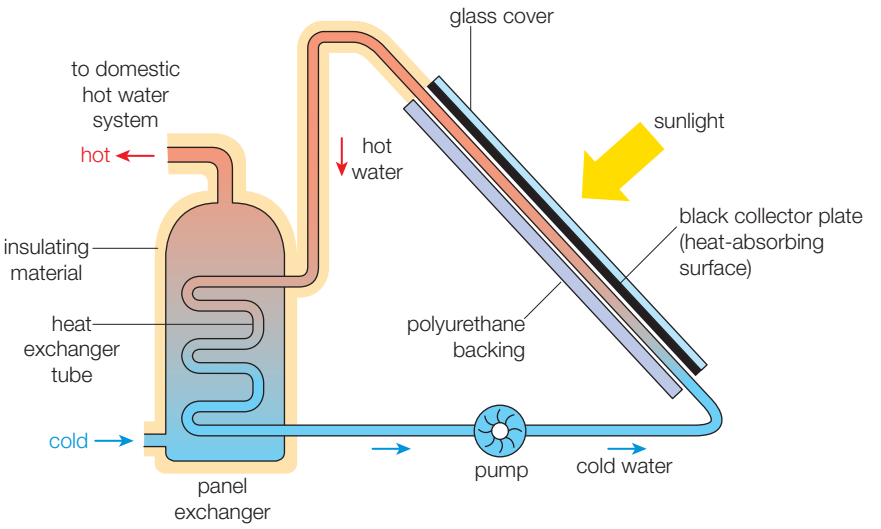


Figure 14.21 Solar water-heater (pump-assisted model).

heat exchanger ›

The heated water leaving the solar panel rises and travels along an insulated tube to the **heat exchanger** in the storage tank. The water flow is assisted by a pump in models in which the storage tank is at or below the level of the collector panel. The cold water entering the storage tank is warmed up at the heat exchanger. The hot-water outlet is placed near the top of the tank because hot water rises within the tank by convection.

Solar panels produce the hottest water on bright, sunny days. The storage tank is well insulated, to keep the water warm for long periods during the night or during times when there is no sunshine.

ITQ9

How could the hot water in the solar heating system shown in figure 14.21 be made to circulate without the aid of a pump? Explain your answer.

Ventilating our buildings in the Caribbean

A problem with most buildings in the sunny Caribbean is that heat tends to flow into such buildings (see red arrows, figure 14.22) for much of the time, making them uncomfortably hot. By comparison, the grey arrows show some heat loss problems encountered in buildings in cold countries in winter (percentages shown are approximate). The diagram shows that the largest percentage of heat flow takes place through the walls and roof of a closed building.

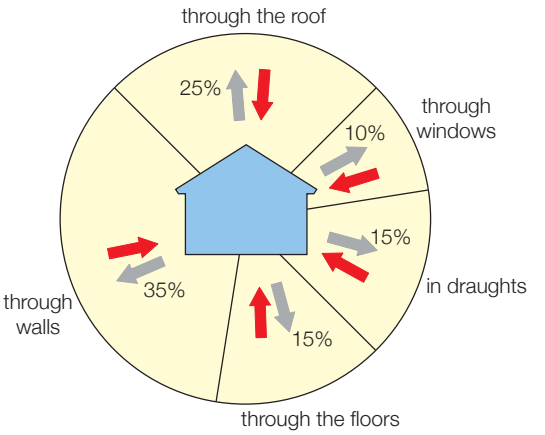


Figure 14.22 Heat exchanges taking place at various parts of a building.

To save on air-conditioning costs in hot weather, we might:

- paint walls white on the outside to reflect the Sun's heat radiation;
- paint roofs white as well, or give them a shiny surface, to reflect the Sun's radiant heat energy;
- make walls of hollow concrete blocks to reduce conduction of heat energy, since the air in the hollow blocks is a poor conductor;

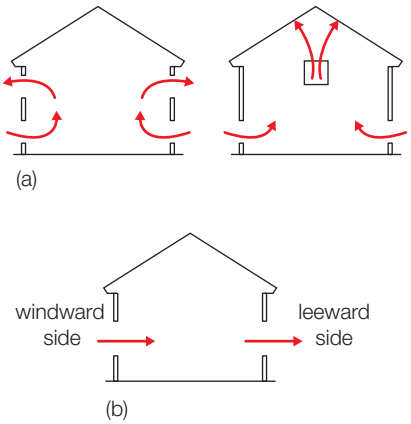


Figure 14.23 Making use of natural ventilation in a building. (a) Convection currents. (b) Sea breeze ventilation.

- use ceilings so that the air in the space between the ceiling and the roof itself will insulate against heat conduction from outside;
- make more use of windows to encourage greater natural heat exchange by convection.

If the building does not have to be kept closed, then a natural ventilation system can be used to help make the building cooler, as shown in figure 14.23 (a). There are openings to the outside at the upper and lower parts of the building. Convection currents are set up, as any heated air will rise and cooler air flows into the building to replace it. Alternatively, as in figure 14.23 (b), open windows on the windward and leeward sides of a building allow natural wind current to remove heat as they flow through the building.

Heat pumps

heat pump

An air-conditioner (figure 13.16) removes heat from the air inside a building by forced convection. Room air is forced over the cooling pipes containing the refrigerant by means of a fan inside the air-conditioner. The refrigerant takes up the heat from the air (see pages 208–209) and subsequently gives up heat to the outside through radiation from the cooling fins of the radiator. If the air-conditioner were 'turned around', it would transfer heat from the outside to the inside of the building. So, a 'turned-around' air-conditioner can be used to heat a room in winter – and at less cost than an electric heater! Modern '**heat pumps**' are used like this, as air-conditioners in summer and as room heaters in winter.

Chapter summary

- Heat energy can travel by conduction, convection or radiation. In the form of radiation it is called radiant heat energy.
- In conduction, heat energy travels along matter, being passed from molecule to molecule in some materials. The material as a whole does not move.
- Metals are good conductors of heat energy because they contain very mobile free electrons that collide with and impart energy to the ionic lattice structure. Non-metals do not have free electrons in their structure. Hence they are poor heat conductors and are called insulators.
- In convection, heat energy is carried by moving particles of matter. Convection takes place in fluids (liquids and gases) since particles in fluids can move around and carry energy along with them.
- Heat energy transfer by radiation does not need a material medium. Radiant energy is transferred by electromagnetic waves
- Black or dark surfaces are good radiators and absorbers of radiant heat energy. Shiny or white surfaces are good reflectors and poor absorbers of radiant heat energy.
- Shorter-wavelength, and to a certain extent, longer-wavelength electromagnetic energy from the Sun can pass through glass (or through carbon dioxide in the atmosphere). Matter that is heated by this energy then emits infrared radiation in its turn. However, this is longer-wavelength radiation, which cannot pass back through the glass (or is trapped by carbon dioxide). Thus the heat energy is trapped in the greenhouse (and by the atmosphere). This is called the 'greenhouse effect'. The greenhouse effect is believed to contribute to global warming because of the increasing amount of carbon dioxide in the Earth's atmosphere.

Answers to ITQs

ITQ1 To help minimise the heat reaching the other side of the bar by means other than conduction along the bar.

ITQ2 The rug is a poorer conductor of heat than the wooden floor. The floor therefore conducts heat away from the warm foot faster than the rug does, so the floor feels colder than the rug, even though both are at the same temperature.

ITQ3 Although snow is cold itself, it is a very poor conductor of heat and is a very good insulator. Thus the heat loss from inside the snow-house to the outside would be small.

ITQ4 The temperature of the air over the land was probably the same as that of the air over the sea, so no convection currents were being formed.

ITQ5 No, because particles in a solid remain in a fixed position relative to each other on account of the strong forces of attraction between them.

ITQ6 A dynamic equilibrium has been set up. The rate at which heat is being lost by the object to the room is equal to the rate at which heat is being gained by the object from the room.

ITQ7 The wall, because it shows a lighter colour. (A concrete wall would absorb much heat during the day, and therefore radiate much, especially during the early part of the night.)

ITQ8 The silvery surface reflects the radiant heat energy coming from the warm room back in the room. So this heat does not reach the liquid.

ITQ9 By placing the storage tank above the collector. Hot water from the collector panel outlet rises and the cooler water at the bottom of the tank would descend to the collector inlet by convection.

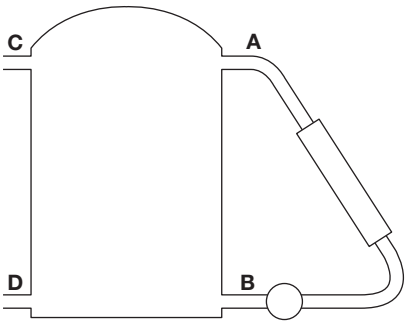
Examination-style questions

1 The outside of most refrigerators is painted white. This helps to reduce heat gain by:

A conduction
B convection
C radiation
D evaporation

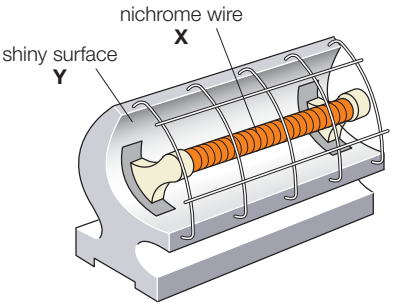
2 Which type of surface is the best radiator of heat energy?

A black
B grey
C silvery
D white

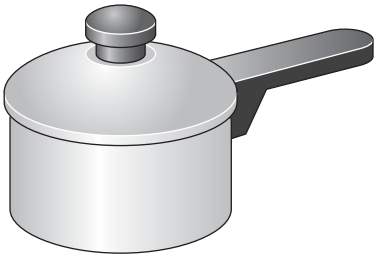


3 The diagram on the left shows a solar water-heating system.
Which part represents the hot-water outlet from the storage tank?

4 The diagram on the right shows an electric room heater.
What is the correct term to use for each of X and Y?



	X	Y
A	radiator	absorber
B	radiator	reflector
C	reflector	radiator
D	reflector	absorber



5 The diagram on the left shows a cooking pot.
Which material might best be used for the handle?

A aluminium
B copper
C polystyrene
D wood

6 When a refrigerator door is opened, the chilled contents of the refrigerator:

A lose cold energy
B lose heat energy
C gain cold energy
D gain heat energy

7 Convection takes place in water because:

A hot water is less dense than cold water
B hot water is more dense than cold water
C water is a good conductor of heat
D water is a poor conductor of heat

8 (i) Describe one way *each* in which conduction, convection and radiation are employed in a domestic refrigerator.

(ii) A 500 W electric refrigerator is controlled by a thermostat so that it is switched on for an average of 6 hours each day.

(a) What is the cost of the operation of the refrigerator in the month of June if one kilowatt-hour of electricity costs 45 cents?

(b) If the overall efficiency of the refrigerator is 40%, how many joules of energy are removed from food in the refrigerator in one day?

(iii) Why do some refrigerator manufacturers advise that the back of a refrigerator be placed at least 10 cm from a wall?

9 On a hot day, an enclosed air-conditioned room is kept at a steady temperature of 25°C when the outside temperature is 30°C. The air-conditioning unit is *effectively* operating at 3000 W (i.e. is removing 3000 J of energy per second from the room).

(i) At what rate is energy entering the room if the temperature is being kept steady?

(ii) The table shows the rates at which energy enters the room from various parts.

Part of room	Rate of gain of energy/W
windows	500
walls	1000
door	
ceiling	800
floor	400

Using the value obtained in part (i), determine:

(a) the rate at which heat enters through the door;

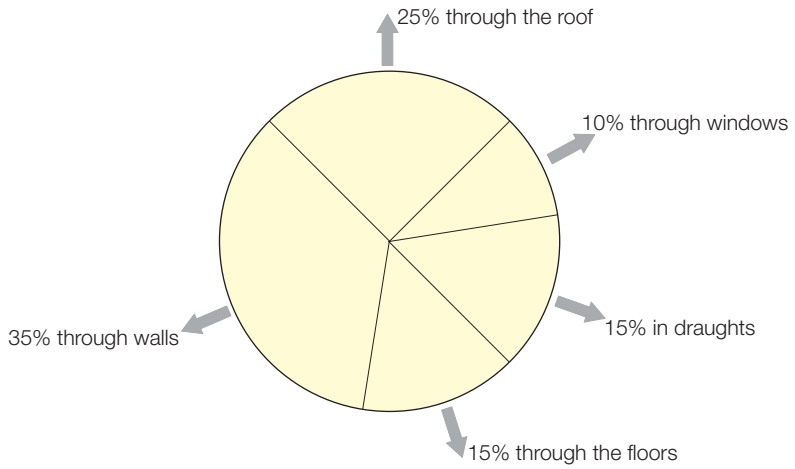
(b) the percentage of heat entering through the walls.

(iii) Suggest *one* way in which heat flow into the room could be reduced:

(a) at the ceiling;

(b) at the windows.

10 The diagram shows the percentage energy losses from a house.



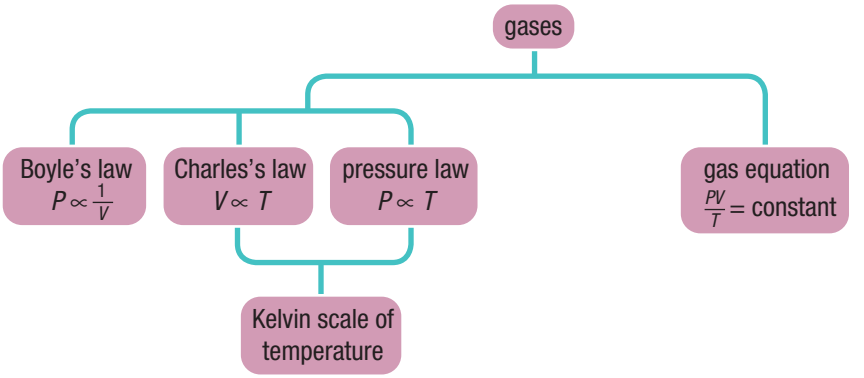
(i) Which *two* parts of the house lose most heat?

(ii) Explain how heat losses in (i) could be reduced through applying principles involved in conduction and radiation.

15 The Behaviour of Gases

By the end of this chapter you should be able to:

- describe experiments to investigate the relationships among the pressure, the volume and the temperature of a gas
- relate graphs of pressure or volume against temperature to the establishment of the Kelvin temperature scale
- use the (approximate) relationship between the Kelvin and Celsius scales, $T = \theta + 273$, where T is in K and θ is in °C
- state the following laws and use them to solve problems:
  - Boyle's law
  - Charles' law
  - the pressure law
- recall that PV/T = constant and use this to solve problems
- give qualitative explanations for the gas laws in terms of the Kinetic Theory



Putting gases to work

CHAPTER 11

internal combustion engine

In chapter 11, we learned that gases exert a pressure on the walls of their containers (page 172). Gases do this through molecular bombardment of the surfaces. Figure 15.1 shows gases at high temperature exerting tremendous pressure on a piston inside the cylinder of a car's internal combustion engine. The pressure forces the piston to move. As the piston moves, it causes parts within the engine to spin.

The above example illustrates a relationship between the pressure of a gas and its temperature. The pressure of a gas is also related to its volume. The ways in which the pressure, volume and temperature of a gas are related to one another will be the focus of this chapter.

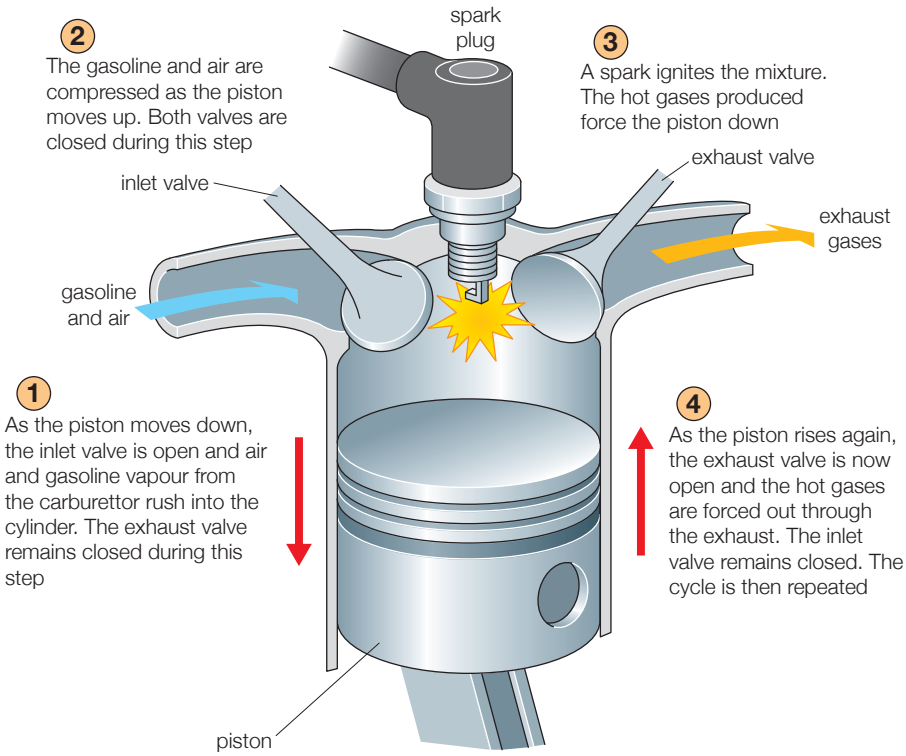


Figure 15.1 Combustion inside a cylinder of a gasoline car engine. The heat produced causes the gases in the cylinder to expand, pushing a piston as they do so.

Boyle's law

CHEMISTRY: gas laws

The relationship between the pressure and volume of a fixed mass of gas was investigated in the 17th century by Robert Boyle, an Irish-born chemist. Boyle found that:

> **The pressure of a fixed mass of gas is inversely proportional to its volume, provided that the temperature of the gas is kept constant.**

Boyle's law ›

This is called **Boyle's law**, and states that, at constant temperature, as the volume, V, of a fixed mass of gas increases, the pressure it exerts, P, decreases by the same factor. Thus, if the volume of the gas is doubled, the pressure of the gas is halved.

Figure 15.2 shows an apparatus that can be used to investigate the variation of pressure, P, with volume, V, for a fixed mass of air. Air is trapped by a column of oil in a sturdy glass tube. The volume, V, of the air can be read directly from a scale by the tube. The column of oil is supplied from the reservoir shown in the diagram. The pressure of the air in the space above the reservoir can be increased by forcing more air into the space, using a foot pump, and this pressure is transmitted through the oil to the air in the glass tube. Thus the pressure, P, of the air in the tube is practically the same as that in the air space above the oil reservoir. This pressure is measured directly with a Bourdon gauge.

It is wise to take the following precautions when the apparatus in figure 15.2 is used to investigate Boyle's law:

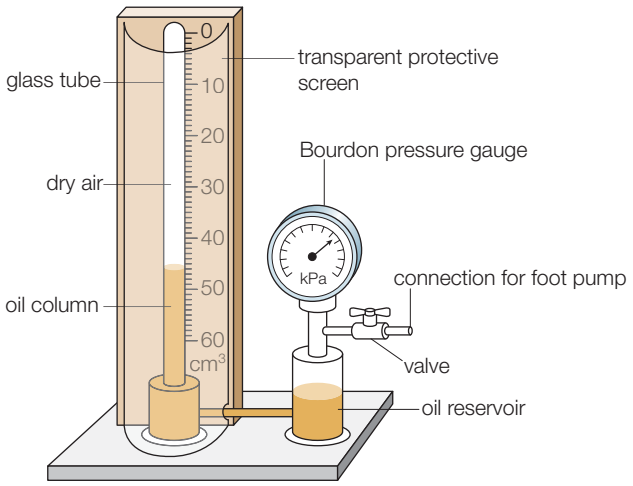


Figure 15.2 Apparatus for investigating the relationship between pressure and volume of a fixed mass of gas.

- A protective transparent plastic screen is placed around the glass tube. This is to safeguard the experimenter in case the tube should shatter when the trapped air is under pressure. Care is also taken not to pressurise the apparatus above the manufacturer's recommended safety ratings.
- Since Boyle's law refers to a fixed mass of gas, the valve is closed and the level of the oil in the tube is observed for about 15 minutes after the pressure has first been increased using the pump. If the level remains steady, it means that there are no leaks in the glass tube. A steady level also means that there are no leaks in the chamber containing the oil. The first reading can now be taken.
- Gases warm up when compressed, and cool down when expanded. After each change in pressure, therefore, the apparatus is left for about 5 minutes to allow the gas to return to room temperature before readings of pressure and volume are taken.

ITQ1 **If a gas of pressure *P* is compressed (e.g. using a piston) to one-third of its original volume, what will be the final pressure exerted by the gas, if the temperature is kept constant?**

ITQ2 **What is the difference between the pressure of the air in the tube and that above the oil in the reservoir in figure 15.2, and why can we ignore it?**

A graph of P against V, using typical experimental data, is shown in figure 15.3. If we check the product PV at a number of points on the graph, we see that, allowing for experimental error, PV is constant. This tells us that the curved graph shows an **inverse relationship** between P and V. Thus, when the volume of the gas is doubled, the gas pressure is halved. We can express this inverse relationship in several ways:

inverse relationship ›

$$V \propto \frac{1}{P}$$

or rearranging

$$P \propto \frac{1}{V}$$

or

$$P = \text{constant} \times \frac{1}{V}$$

or, rearranging this equation again,

$$PV = \text{constant}$$

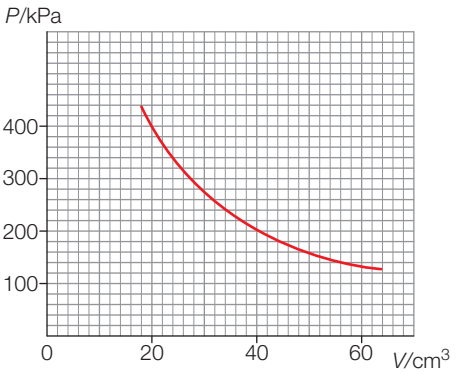


Figure 15.3 Graph showing the inverse relationship between the pressure, *P*, of a gas and its volume, *V*.

Bourdon gauges measure the actual pressure of a gas. However, tyre pressure gauges measure how much the air pressure in a tyre exceeds atmospheric pressure. Thus, the actual air pressure in a tyre = pressure shown on the tyre gauge + atmospheric pressure.

MATHEMATICS: inverse relationships

MATHEMATICS: direct relationship

The graph of P against $1/V$ in figure 15.4 is a straight line through the origin. This also shows that P is **directly proportional** to $1/V$ (P = constant $\times$ $1/V$), and hence that P varies inversely with V ($P \propto 1/V$ and $V \propto 1/P$).

directly proportional ›

If PV = constant, then, at constant temperature,

$$P_1V_1 = P_2V_2$$

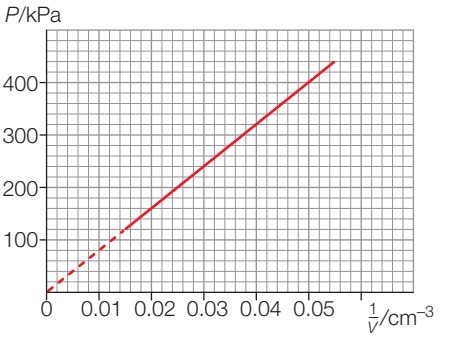


Figure 15.4 Graph showing direct relationship between *P* and $1/V$ for a fixed mass of gas at a constant temperature.

where P_1 is the gas pressure when the volume is V_1 and P_2 is the gas pressure when the volume is V_2.

ITQ3 **According to the graph in figure 15.4, what would be the volume of the gas at a pressure of 200 kPa?**

Practical activity 15.1

Investigating Boyle's law (using Boyle's Law apparatus)



Caution: Use safety goggles



Caution: Risk of explosion

Using apparatus similar to figure 15.2, and referring to procedure and precautions above, do the following:

1 Obtain and record a variety of pressures, P, and corresponding volumes, V, for the gas trapped in the tube.

2 Plot graphs of P vs. V; P vs. $1/V$; and PV vs. P.

Comment on whether each of the graphs was in accordance with Boyle's law.

Practical activity 15.2

Investigating Boyle's law (using simple apparatus)



Caution: Use safety goggles



Caution: Risk of explosion

You will need:

- 20 cm^3 graduated gas syringe with sealed tip on nozzle
- 4 kg (40 N) top-pan spring balance
- Ruler graduated in mm.

Method

1 Slacken the nozzle and pull the plunger to allow air to enter the barrel to a volume of about 20 cm^3.

2 Tighten the nozzle (very tightly indeed) so as to trap the air in the barrel.

3 Test the syringe for leaks. This is done by pressing and then releasing the plunger with the gas exit properly sealed. If the plunger returns to its original position when released, it means that the barrel is airtight.

4 Measure the length, l, of the barrel of the syringe.

5 Using the value obtained for l, and the corresponding volume of the barrel, determine the cross-sectional area, A, of the piston.

6 Invert the piston on to the balance as shown in figure 15.5 and set the balance to zero. Note the volume V_0.

7 Holding the flange, press down on the syringe until the balance reads 0.4 kg (4 N). Note the volume, V, of the air in the barrel. Calculate the equivalent pressure, P (P = force/A) being exerted on the gas due to this applied force.

8 Repeat step 7 for different values on the balance, up to 2.0 kg (20 N) so as to determine a range of values for P and corresponding values for V. (The readings should be carried out as quickly as possible in case minor leaks should develop in the equipment.)

9 After the last measurement has been taken, re-check the volume of air in the barrel when there is no applied force.

10 Add P_0, the value of the atmospheric pressure, to each value of P, to obtain a range of values, P_T, of the total pressure acting on the air in the barrel.

11 Plot a graph of P_T against $1/V$.

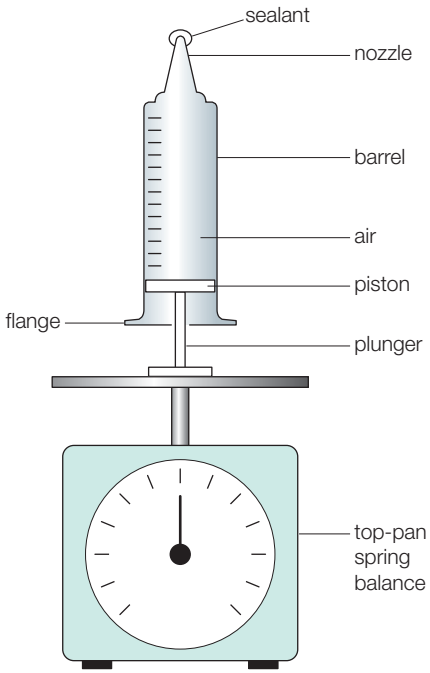


Figure 15.5 Investigating Boyle's law using simple apparatus.

Questions

1 (a) What *two* features would you expect your graph to show if Boyle's law was verified by the experiment?

(b) Did your graph indicate that Boyle's law was verified?

2 Why was it recommended that you hold the flange rather than the barrel when pressing the gas syringe against the balance?

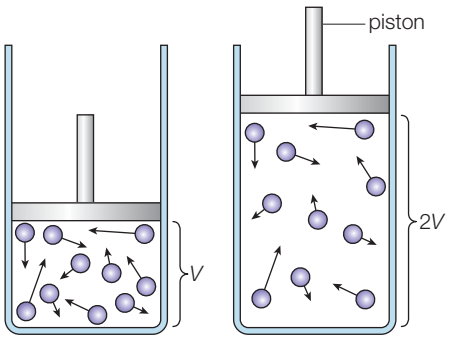
3 Why is it necessary to carry out step 9?

4 How did you determine the atmospheric pressure, P_0?

Extension

Find out how Boyle's law can be used to explain the mechanism of breathing in humans. In what *two* ways is the volume change brought about?

BIOLOGY: breathing in mammals



$V \propto \frac{1}{P}$ at constant temperature

Figure 15.6 If the volume of a gas is doubled, by moving a piston as shown, the molecules of the gas have fewer collisions with the container walls in a given time, and therefore the pressure is reduced.

Worked example 15.1: Boyle's law

80 cm³ of gas, originally at a pressure of 100 kPa, is compressed to 60 cm³. What is the final pressure of the gas?

Solution

Initial pressure, $P_1 = 100\,\text{kPa}$

Initial volume, $V_1 = 80\,\text{cm}^3$

Final pressure, $P_2 = ?$

Final volume, $V_2 = 60\,\text{cm}^3$

Assuming Boyle's law holds, then

$$P_1V_1 = P_2V_2$$

$$P_2 = \frac{P_1V_1}{V_2}$$

$$P_2 = \frac{100\,\text{kPa} \times 80\,\text{cm}^3}{60\,\text{cm}^3}$$

$$P_2 = 130\,\text{kPa} \text{ (2 sig. fig.)}$$

CHAPTER 11

ITQ4

Why do the bubbles of air from a diver's mask (figure 15.7) expand as they rise through the water?



Figure 15.7 Bubbles of air expand as they rise from a diver's mask.

The Kinetic Theory and Boyle's law

In chapter 11 we learned that gases exert a pressure on a surface through molecular bombardment (page 172). If the volume of a fixed mass of gas in a container is increased, fewer molecules will bombard the walls each second since it would take, on average, a longer time for the molecules to traverse the enlarged space within the container (figure 15.6). Hence the pressure of the gas is reduced.

Although Boyle's law holds for all gases, it is not applicable at very high gas pressures. This is because, at high pressures, the molecules are much closer together and the volume of the particles themselves becomes significant compared with the overall volume of the gas. The particles tend to collide with each other more frequently, and cannot move as freely to collide with the walls of the container. In addition, at closer distances, molecules attract each other strongly. This attraction has the effect of reducing the speeds with which gas molecules collide with the walls of the container. These three factors lead to a reduction in the pressure of the gas on the walls.

Charles' law

CHEMISTRY: gas laws

The French scientist, Jacques Charles, discovered in 1787 that:

The volume of a fixed mass of gas is directly proportional to its absolute temperature, provided that the pressure of the gas is kept constant.

Charles' law

This statement is known as **Charles' law**, and may be verified using the apparatus shown in figure 15.8. A fixed mass of air is trapped in a capillary tube by means of a drop of concentrated sulphuric acid. The trapped air, at atmospheric pressure (assumed constant for the duration of the experiment), is heated in a water bath. The length of the air column, as a measure of the volume of the air, is noted at different temperatures.

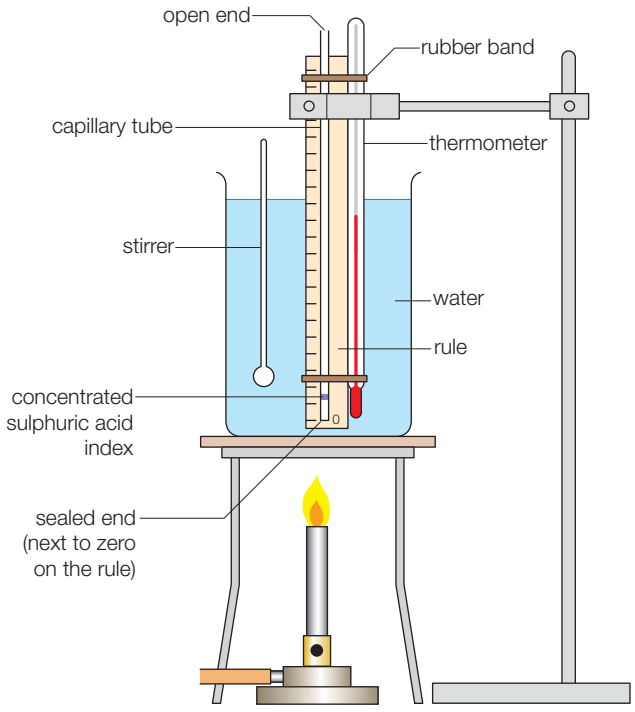


Figure 15.8 Apparatus for investigating the relationship between volume and temperature for a fixed mass of gas at constant pressure.

Safety notes: The drop of concentrated sulphuric acid also helps to keep the air dry and is therefore better than mercury. However, concentrated sulphuric acid is corrosive to the skin, mercury vapour is toxic and glass is fragile. You should therefore be very careful.

The water is stirred before each temperature reading, since warm water tends to settle above colder water.

A graph of volume, V, against temperature, θ (in °C), using typical experimental data, is a straight line that does not pass through the origin (figure 15.9). The straight line indicates that the air expands uniformly with temperature, i.e. that equal changes in temperature result in equal changes in volume.

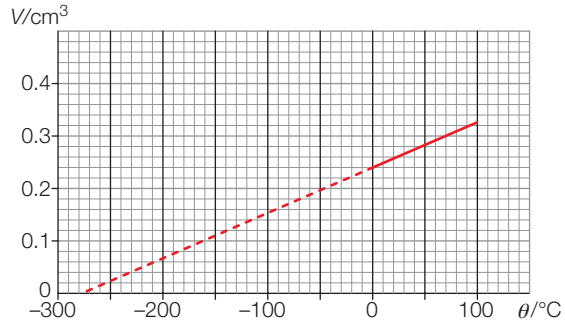


Figure 15.9 Graph of volume, V, against temperature, θ, for a fixed mass of gas at constant pressure.

absolute zero

If the graph is continued to the left (broken line) it cuts the temperature axis at about –273°C. This temperature is believed to be the lowest attainable temperature or the **absolute zero** of temperature. The absolute scale, or Kelvin scale, uses this temperature as its zero (see chapter 12, pages 186–188). Temperatures on this scale are measured in kelvin (K), where an interval (temperature change) of one kelvin is equal to an interval of one Celsius degree. The conversion between absolute temperature, T, and Celsius temperature, θ, is therefore given by

$$T = \theta + 273 \quad \text{(where } T \text{ is in } K \text{ and } \theta \text{ is in °C)}$$

A graph of V against T is a straight line passing through the origin. This means that

$$V \propto T$$

or

$$V = \text{constant} \times T$$

IT Q5

Suggest a safety precaution that should be taken when using the apparatus shown for verifying Charles' law (figure 15.8).

IT Q6

Room temperature is 27°C. What temperature is this on the absolute (Kelvin) scale?

Rearranging the equation, we get

$$\frac{V}{T} = \text{constant}$$

MATHEMATICS: direct proportion; linear relationship

Hence, using the usual notation,

$$\frac{V_1}{T_1} = \frac{V_2}{T_2}$$

The above three equations are all expressions of Charles' law.

Worked example 15.2: A balloon in the freezer

At room temperature (27°C) the volume of air trapped in a balloon is 900 cm³. What would the volume of the air be if the balloon was cooled to –10°C in a freezer?

Solution

(Note: To apply Charles' law, we must first convert temperatures to kelvin.)

Initial temperature, $T_1 = 27 + 273 = 300\,\text{K}$

Final temperature, $T_2 = -10 + 273 = 263\,\text{K}$

Initial volume, $V_1 = 900\,\text{cm}^3$

Final volume, $V_2 = ?$

Assuming Charles' law holds,

$$\frac{V_1}{T_1} = \frac{V_2}{T_2}$$

Rearranging,

$$V_2 = \frac{V_1 T_2}{T_1}$$

$$= \frac{900\,\text{cm}^3 \times 263\,\text{K}}{300\,\text{K}}$$

$$V_2 = 790\,\text{cm}^3 \text{ (2 sig. fig.)}$$

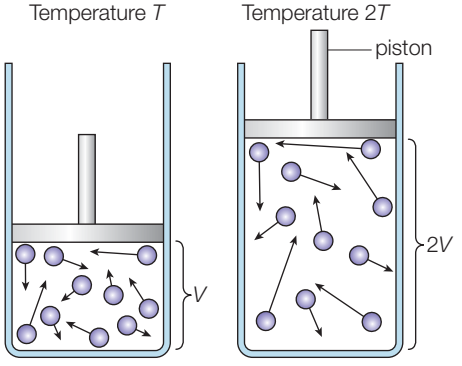


Figure 15.10 As the gas is heated, the average speed of the molecules is increased.

The Kinetic Theory and Charles' law

When the temperature of a gas is increased, the average speed of the molecules increases (see page 172). Hence, if the temperature of the gas in figure 15.10 is increased, there will be more frequent collisions with the piston, which would increase the gas pressure. If the pressure on the piston is kept constant, the piston will therefore have to rise and the volume will increase, until the gas pressure is equal once more to the constant pressure exerted by the piston.

At very low temperatures, Charles' law may not apply, since below certain temperatures gases become liquid.

The pressure law

As you might expect, there is also a relationship between the pressure and the temperature of a given volume of gas. This relationship is expressed in the following pressure law:

CHEMISTRY: gas laws

The pressure, *P*, of a fixed mass of gas is directly proportional to its absolute temperature, *T*, provided that its volume is constant.

Figure 15.11 shows an apparatus that can be used to verify the pressure law. Data from such experiments give pressure–temperature graphs like the one in figure 15.12. This shows similarities with the volume–temperature graph of figure 15.9.

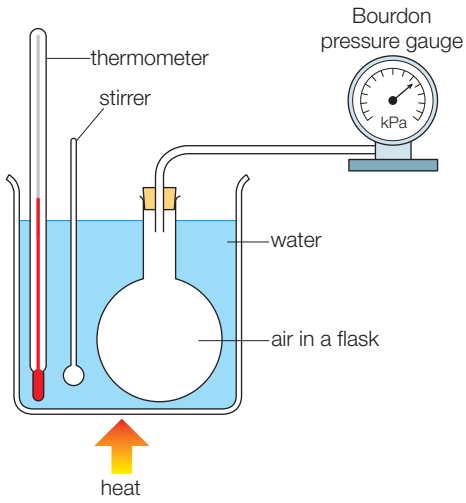


Figure 15.11 Apparatus for verifying the pressure law.

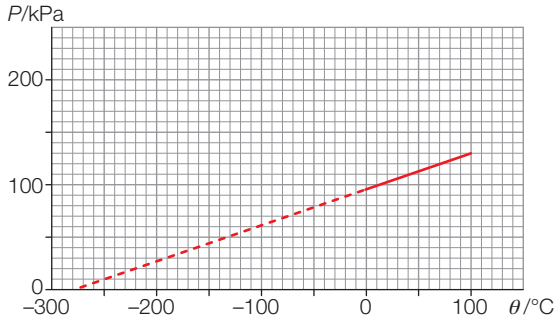


Figure 15.12 Graph of pressure against temperature for a fixed mass of gas kept at constant volume.

ITQ7
Give *two* similarities between the volume–temperature graph of figure 15.9 and the pressure–temperature graph of figure 15.12.

The pressure–temperature graph, when extended, cuts the temperature axis at approximately –273°C, just like the volume–temperature graph. This indicates, once again, that there is an absolute zero of temperature at –273°C.

If the temperatures in figure 15.12 are shown in kelvins, then the graph is a straight line passing through the origin. This means that

$$P \propto T$$

or

$$P = \text{constant} \times T$$

MATHEMATICS: direct proportion; linear relationship

Rearranging the equation,

$$\frac{P}{T} = \text{constant}$$

Hence, using the usual notation,

$$\frac{P_1}{T_1} = \frac{P_2}{T_2}$$



Figure 15.13 At the top of Mount Kilimanjaro, in the tropics, there is snow!

ITQ8
Mount Kilimanjaro (figure 15.13) is very near to the equator. Hot air rises towards mountain tops, so how is it that there is snow on the top of Mount Kilimanjaro?

A type of thermometer, the constant-volume gas thermometer, makes use of the linear pressure–temperature relationship, with absolute zero as the zero of its scale. This thermometer is used in scientific work where accurate measurements of temperature are required. Figure 15.11 therefore represents a simple type of constant-volume air thermometer.

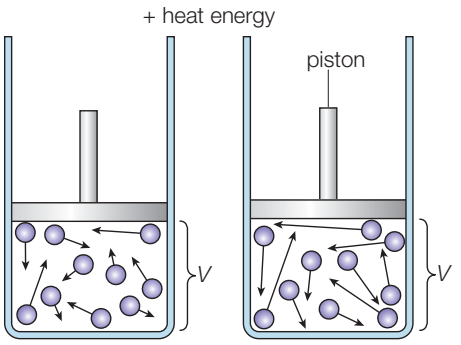


Figure 15.14 The temperature of the gas is increased and the volume is kept constant. The increased molecular speeds result in more frequent molecular bombardment on the piston, and hence greater pressure.

The Kinetic Theory and the pressure law

According to the Kinetic Theory, as the gas shown in figure 15.14 is heated, increased molecular speeds lead to a greater and more frequent molecular bombardment of the piston. If the volume is not allowed to increase, then a greater pressure must be applied to the piston to counter the greater pressure exerted by the gas. At very low temperatures or at very high pressures, however, the pressure law does not apply since the molecules of the gas will be very close to each other and exert forces of attraction between them.

Worked example 15.3: Car tyre pressure

A car tyre is inflated to a gauge pressure (see the note on page 228) of 30 psi in the early morning when the temperature is 25°C. What will the gauge pressure be at midday when the tyre temperature has reached 40°C? (Assume atmospheric pressure = 15 psi.)

Solution

(Note: To apply the pressure law, we must first convert temperatures to kelvins and convert tyre gauge pressure to the pressure on the gas.)

Initial pressure of the air in the tyre, $P_1 = 30\,\text{psi} + 15\,\text{psi} = 45\,\text{psi}$

Initial temperature, $T_1 = 25 + 273 = 298\,\text{K}$

Final temperature, $T_2 = 40 + 273 = 313\,\text{K}$

Final pressure of the air in the tyre, $P_2 = ?$

Assuming the pressure law holds

$$\frac{P_1}{T_1} = \frac{P_2}{T_2}$$

$$P_2 = \frac{P_1 T_2}{T_1}$$

$$P_2 = \frac{45 \times 313}{298}$$

$$P_2 = 47\,\text{psi}\ (2\,\text{sig. fig.})$$

Final gauge pressure = 47 – 15 = 32 psi

The psi (pounds per square inch) is a pressure unit still used on some car tyre pressure gauges. One psi is approximately equal to 6900 N m^{-2} (or Pa).

Tyre manufacturers usually recommend that tyre pressure checks be done early in the morning. If tyres are inflated to the recommended pressures in the heat of the day, the tyres may become under-inflated as the temperature, and hence the pressure, drops late in the afternoon.

The general gas equation

CHEMISTRY: gas laws
gas equation ›

The equations for Boyle's law (PV = constant) and Charles' law (V/T = constant) may be combined into a single equation called the general **gas equation**:

$$\frac{PV}{T} = \text{constant}$$

When we use the gas equation, the temperatures must always be expressed in kelvin.

We often use this equation to solve problems where the pressure, volume and temperature of a gas all change at the same time. If the values change from P_1, V_1 and T_1 to P_2, V_2 and T_2, and the mass of the gas is fixed, we can write:

$$\frac{P_1 V_1}{T_1} = \frac{P_2 V_2}{T_2}$$

This equation is very useful since, in practice, gases usually undergo changes in pressure, volume and temperature all at once.



Figure 15.15 A partially inflated weather balloon. Radio-transmitting weather devices, called radiosondes (not shown), are attached to the bottom to be hoisted into the atmosphere. As the balloon rises, the volume increases.

Internet search terms: U-tube gas laws/ virtual experiments gas laws

Worked example 15.4: The weather balloon

A weather balloon (figure 15.15) is partially inflated with helium gas to a volume of 2.0 m³ at sea level, where the pressure is 101 kPa, and at a temperature of 27°C. What would be the volume of the balloon at a height of 9 km, where the atmospheric pressure is 40 kPa and the air temperature is 10°C?

Solution

Initial pressure, $P_1 = 101\,\text{kPa}$

Initial volume, $V_1 = 2.0\,\text{m}^3$

Initial temperature, $T_1 = 27 + 273 = 399\,\text{K}$

Final pressure, $P_2 = 40\,\text{kPa}$

Final volume, $V_2 = ?$

Final temperature, $T_2 = 10 + 273 = 283\,\text{K}$

Using the gas equation

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2}$$

Rearranging,

$$V_2 = \frac{P_1V_1}{T_1} = \frac{T_2}{P_2}$$

$$V_2 = \frac{101\,\text{kPa} \times 2.0\,\text{m}^3 \times 283\,\text{K}}{300\,\text{K} \times 40\,\text{kPa}}$$

$$= 4.8\,\text{m}^3 \text{ (2 sig. fig.)}$$

Chapter summary

- Boyle's law states that the pressure, *P*, of a fixed mass of gas is inversely proportional to its volume, *V*, provided the temperature remains constant. The equation
  $PV = \text{constant}$
  expresses this relationship. Boyle's law does not hold, however, at very high pressures.
- Charles' law states that the volume, *V*, of a fixed mass of gas is directly proportional to its absolute temperature, *T*, provided that the pressure of the gas is kept constant. The equation
  $\frac{V}{T} = \text{constant}$
  expresses this relationship. Charles' law does not hold at very low temperatures.
- The pressure law states that the pressure, *P*, of a fixed mass of gas is directly proportional to its absolute temperature, *T*, provided that the volume is kept constant. The equation
  $\frac{P}{T} = \text{constant}$
  expresses this relationship. The pressure law does not hold at very low temperatures or at very high pressures.
- Boyle's law and Charles' law may be combined to give the equation
  $\frac{PV}{T} = \text{constant}$
  This is known as the general gas equation.

Answers to ITQs

ITQ1 $3P$

ITQ2 The pressure P_o above the oil reservoir in the gauge is equal to the pressure P in the glass tube plus the pressure due to the weight of the oil column, i.e. $P + h\rho g$, where h is the difference between the height of the oil column and the height of the oil reservoir, l is the density of the oil and g is the acceleration due to gravity (see chapter 10, page 150). The term $h\rho g$ is small compared with P_o, so $P_o \approx P$.

CHAPTER 10

ITQ3 (Since $\frac{1}{V} = 0.025\,\text{cm}^{-3}$) $40\,\text{cm}^3$

ITQ4 The pressure of the air in the bubble is the same as that of the surrounding water. This pressure is lower nearer the surface of the water. Hence, following Boyle's law, the volume of the air bubbles increases as they move from a region of higher pressure (at the diver's depth) to one of lower pressure (near the surface).

ITQ5 Keep flammable substances, loose clothing and long hair away from the Bunsen flame. (**Caution:** Take extreme care when handling concentrated sulphuric acid.)

ITQ6 $T\,\text{K} = 27 + 273 = 300\,\text{K}$

ITQ7 They are both linear. They both intercept the Celsius temperature axis at approximately –273°C.

ITQ8 Higher up, the atmospheric pressure is less than it is at sea level. Hence, as an air mass rises, its pressure drops and therefore its temperature, according to the pressure law, falls. (Strictly speaking, the pressure law holds only for gases that are totally enclosed. The air mass in question here is partially enclosed as it is forced up the mountainsides.)

Examination-style questions

Items 1 and 2

Choose the *best* answer A, B, C or D that applies to each item.

A Boyle's law
B Charles' law
C Pressure law
D Gas equation

1 An 'air' thermometer can be made by connecting a Bourdon pressure gauge to a flask containing air.

2 An 'air' thermometer can be made by using a sulphuric acid 'index' to trap air in a capillary tube sealed at one end.

3 According to Charles' law, when the absolute temperature of a fixed mass of gas at constant pressure is doubled, the volume of the gas:

A is halved
B remains the same
C is doubled
D is quadrupled

4 For the pressure of a gas at 27.0°C to double whilst its volume is kept constant, the temperature must be changed to:

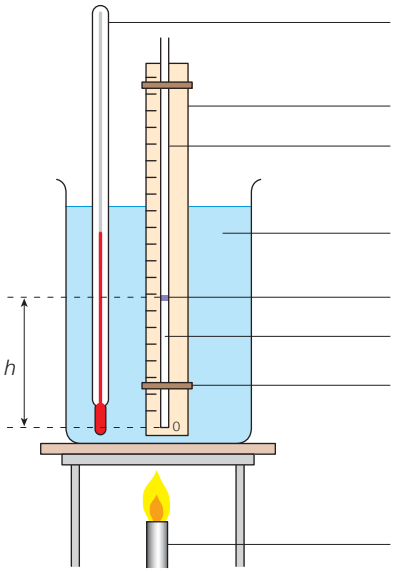
A 13.5°C
B 54.0°C
C 327°C
D 600°C

5 (i) State Boyle's law.

(ii) State Charles' law.

6 A sample of gas is held in a 2.7 m^3 volume at a pressure of 226 kPa. The temperature is kept constant while the pressure is increased to 484 kPa. Calculate the new volume of the gas.

7 The diagram on the right shows an apparatus that can be used to investigate how the volume of a fixed mass of gas varies with its temperature at constant pressure.



(i) Label the diagram and give it a title.

(ii) How is constant pressure achieved in the set-up above?

(iii) An item that is quite useful when taking temperature readings in water is missing in the diagram.

(a) Name the missing item.

(b) State the purpose of the missing item.

(iv) The height, *h*, can be used as a measure of the volume, *V*, of the gas.

Give a reason for this.

(v) A student converted the heights recorded, *h*, into the corresponding volumes, *V*. The student obtained the following values of *V* for values of temperature, θ.

V/mm^3	21.4	22.6	24	25.2	26.6
θ/°C	20	35	50	70	90

(a) Draw a graph of volume, *V* (*y*-axis), against temperature, θ (*x*-axis). The temperature axis should go from –300°C to +100°C.

(b) From the graph, determine the volume of gas at 0°C and the temperature of the gas when its volume is 0 mm^3.

(vi) Calculate the gradient of the graph.

(vii) State *two* sources of error in this experiment.

8 A 20-litre sample of argon gas at a temperature of 0°C is at an atmospheric pressure of 101 kPa. The temperature is lowered to the boiling point of nitrogen, –196°C, while the pressure is increased to 145 kPa. Find the new volume of the argon sample. (Take the temperature of absolute zero to be –273°C.)

9 (i) State Boyle's law.

(ii) (a) State *two* experimental conditions that must be met when performing an experiment to investigate Boyle's law.

(b) Describe how the experimental conditions in (b) (i) are met when using air enclosed in a glass tube that is connected to a Bourdon pressure gauge.

(c) Describe *one* safety precaution to be taken in such an investigation.

(iii) The following results were obtained in an experiment to investigate Boyle's law.

Pressure, *P*/kPa	100	160	200	240	300
Volume, *V*/cm^3	83	50	40	33	27

(a) Using the above results, plot a graph of *PV* (*y*-axis) against *P* (*x*-axis).

(b) Comment on the extent to which your graph shows that the results are in agreement with Boyle's law.

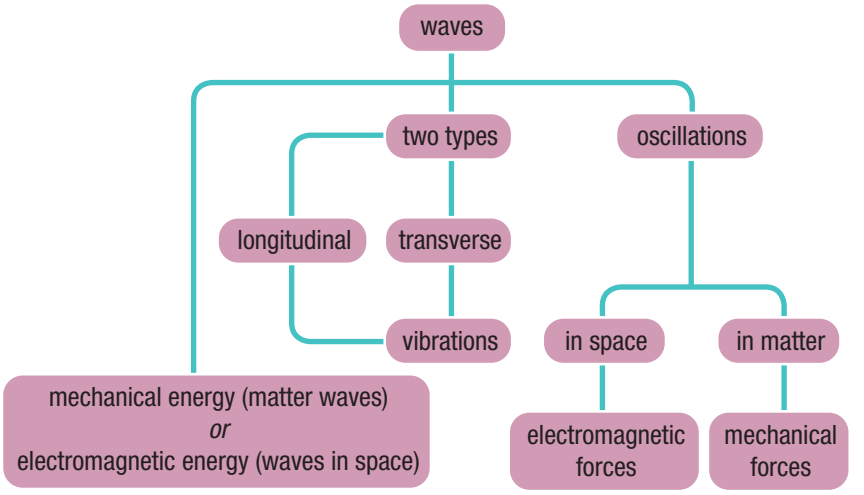
Section C:

Waves and Light

16 What Are Waves and How Are They Produced?

By the end of this chapter you should be able to:

- recall the nature of a vibration or an oscillation
- understand how a pulse is produced by an oscillating object
- appreciate that a wave is a set of continuous pulses
- describe the production of waves using springs and ripple tanks
- distinguish between the two basic types of wave – transverse and longitudinal
- recall the reason why some transverse waves and some longitudinal waves are described as progressive waves or travelling waves
- recall that all waves transmit energy originating at their source
- recall that some waves (called matter waves) need a medium (matter) for propagation, while others (electromagnetic waves) do not and can travel in a vacuum
- recall that all longitudinal waves need a medium for propagation
- draw diagrams to represent transverse waves
- draw diagrams to represent longitudinal waves
- represent transverse waves and longitudinal waves on displacement–position graphs
- represent transverse waves and longitudinal waves on displacement–time graphs
- use straight lines to represent the direction of energy flow (rays)



Introduction

Although we may not realise it, waves of one kind or another are all around us. We can hear because sound waves travel through the air and strike our eardrums. We see because light waves reach our eyes. Our intercontinental telecommunication systems depend on waves that travel through the upper atmosphere, and the ripples on the seashore are also waves. Indeed, the heat we feel from the Sun, or when we stand in front of a glowing fire, is brought

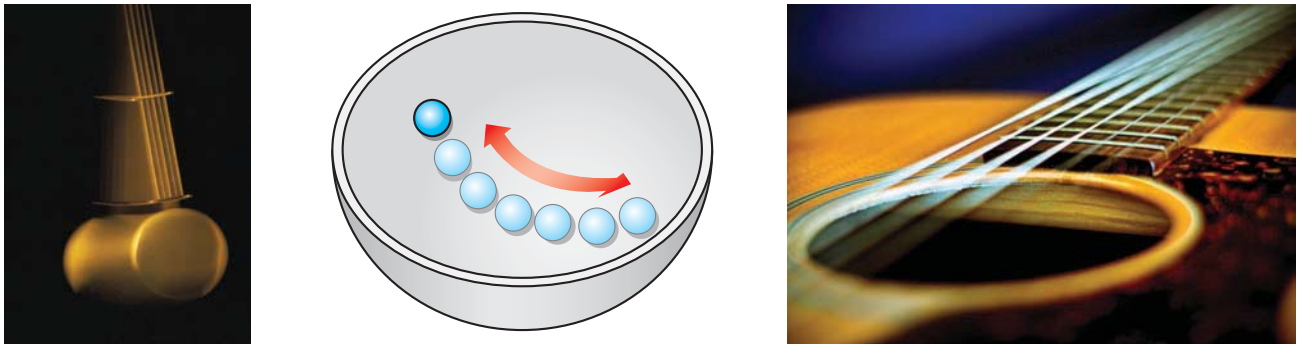
to us by waves. So what is a **wave**? Put simply, a wave could be defined as a disturbance which moves (generally) along a medium repeating itself in any one position in that medium at regular intervals.

Some waves, such as sound waves, require a material medium in order to reach us, while others, like the microwaves used for telecommunications, or radio waves and the heat and light from the Sun, can travel through space. All of them, however, have something in common: vibrations are involved in the transmission of waves. This chapter will deal with vibrations and the ways in which two basic types of wave can be produced. We will consider the nature of waves and examine how they are made up and some of their properties.

How are waves produced?

oscillation vibration

In order to produce a wave, we must first have some sort of oscillation or vibration. An **oscillation** or **vibration** is a motion in which a body moves to and fro about a fixed point. For example, in the swinging pendulum shown in figure 16.1 (a), the bob moves to and fro, about the lowest point of its path; we would call its motion a vibration or an oscillation. There are many other familiar examples of oscillation, for example: a marble rolling up and down the sides of a bowl (figure 16.1 (b)), the swaying of a coconut tree in a strong breeze, what we feel when a mild earth tremor occurs, and the way a stretched string or wire behaves when plucked (figure 16.1 (c)).



(a) (b) (c)

Figure 16.1 Examples of oscillations: (a) a swinging pendulum bob; (b) a marble rolling in a bowl; (c) a stretched string when it is plucked.

periodic

Oscillations that are repeated at regular intervals are described as '**periodic**'.

To produce waves in a material, such as a body of water or a slinky, we would need to have an oscillation that transfers its energy to the water or to the slinky.

One way of producing a wave in a slinky is to:

- stretch the slinky on a long smooth surface (such as a laboratory bench-top), with one end of the slinky held fixed;
- attach a vibrating body (a person's hand moving from side to side will do) to its other end.

pulse

To produce a wave on the slinky, the hand is moved regularly to and fro across the bench at right angles to the length of the slinky. If the hand is moved to and fro only once, we will have a slight disturbance on the slinky. This disturbance is called a **pulse**. See a diagram of one in figure 16.2 (a). This pulse is obtained if the hand is moved once from the rest position, O, to the left at A and back to its starting point, O. The pulse travels along the slinky away from the moving hand and towards the fixed end, and so is called a travelling pulse.

If the hand is moved from the rest position to the left and back past its starting position to the right, B, and back to O, the pulse has a different shape

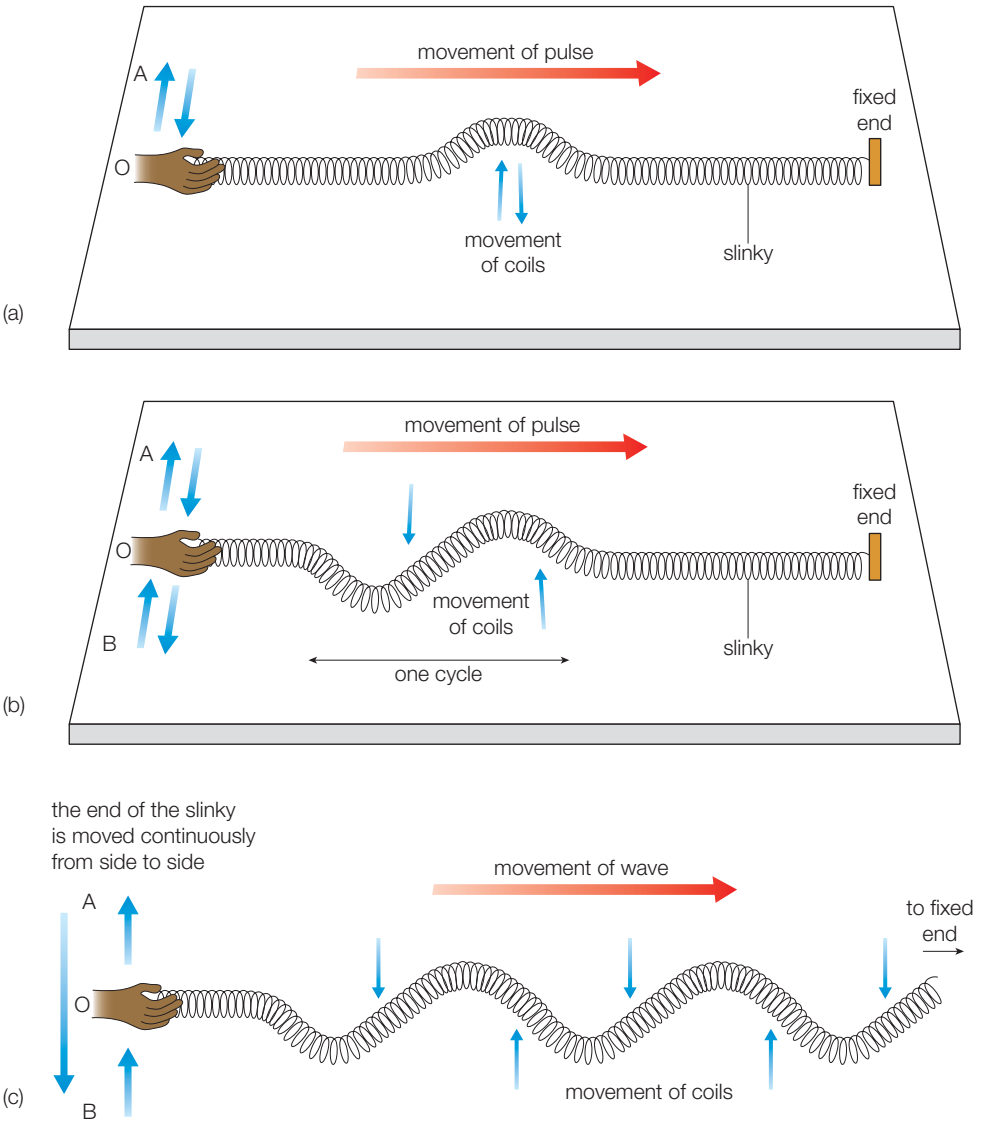


Figure 16.2 (a) A pulse in a slinky. (b) A single wave cycle. (c) A wave travelling along the slinky.

ITQ1

With reference to the diagram of figure 16.2 (b), what happens to the energy of the hand when it stops moving, having done one complete oscillation?

cycle ›

– it is now one **cycle** of a wave, since we will have two pulses, one to the left and one to the right, as shown in figure 16.2 (b). These two pulses together give one cycle of a wave

If the hand is moved continuously from side to side about the starting position, O, we see a set of pulses travelling in succession along the slinky and towards the fixed end. This continuous set of pulses (like the one in figure 16.2 (b)) travelling from the source (the oscillator or the vibrator), along the slinky (the medium), is a wave. Since the pulses travel away from the hand generating them, the wave is called a **travelling wave** (or a **progressive wave**) (see figure 16.2 (c)).

travelling wave (or progressive wave) ›

crests ›

troughs ›

The points on the wave that are furthest to the left we might call **crests** (highest points) and those points furthest to the right we might call **troughs** (lowest points). If you look carefully at those points we call crests and troughs, we notice that they are moving continuously towards the far end of the slinky. In a travelling or progressive wave of this kind, therefore, crests and troughs continuously move from the vibrating object, the source, outwards and away from it. This is a common feature of all progressive or travelling waves of this type – crests and troughs move outwards away from the source along the medium.

matter wave ›

A wave travelling in a medium transfers energy through that medium.

Here, the moving hand is the vibrating source and the slinky is the medium (material) along which the wave travels. This wave is an example of a **matter wave**, so-called because the wave is travelling along a material (matter).

As long as the oscillator continues to oscillate, pulses will continue to leave the source and move along the slinky. Put differently, as long as the vibrating source continues to expend energy by moving to and fro, waves will continue to be generated, taking energy with them from the source. We know that the medium takes energy since we can see the turns of the slinky moving.

When a wave travels through a medium, the medium as a whole does not change its position.

Note carefully that when we say that the wave 'travels', what we mean is that the 'disturbances' (or the pulses) with their crests and troughs move along the slinky – the slinky as a whole does not move.

Transverse and longitudinal waves

Transverse waves

In the waves described above, there is no movement of any part of the slinky along the length of the slinky itself. If you tie a short length of brightly coloured string to a point on the slinky and watch it closely as the wave travels past, you will see that the string moves only at right angles to the length of the slinky.

Because the movement of points on the slinky is perpendicular to the direction of travel of the wave (which is parallel to the length of the slinky), the wave produced is called a **transverse wave** ('trans' means 'across'). Figure 16.3 shows a travelling (or progressive) transverse wave.

transverse wave ›

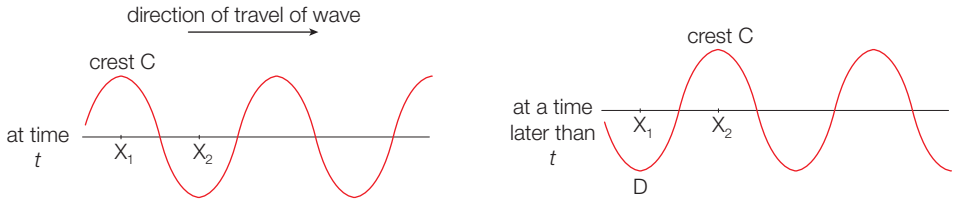


Figure 16.3 Every point on a transverse wave vibrates at right angles to the wave direction. At time *t* the crest C is at X_1. A little later than *t*, there is now a trough D at X_1, and crest C seems to have moved to the right to X_2. We say that the wave has travelled from X_1 to X_2; this wave is therefore a travelling (or progressive) wave.

There are many examples of transverse waves in Nature. Some are matter waves and some are not. Examples of matter waves are the ripples that move in water and the waves on a slinky. The transverse waves that bring heat and light to the Earth from the Sun through space do not need any matter for propagation. They are **electromagnetic waves**, so-called because the energy they carry is electrical and magnetic in nature. These waves are also transverse. There will be further discussions about electromagnetic waves in chapter 19.

electromagnetic wave ›

CHAPTER 19

Longitudinal waves

Longitudinal means 'running lengthwise'.

We can use a slinky to obtain longitudinal waves. But what are longitudinal waves? How do they differ from transverse waves?

We may obtain a longitudinal wave as follows. Stretch the slinky on a long laboratory bench with one end of the slinky fixed, as before. Move the other end of the slinky forwards about 30 cm and back to its starting point once, thus alternately shortening and extending the slinky (figure 16.4 (a)). You will see a pulse move along the slinky away from your hand. If you look carefully as you do this, you will notice that where the pulse passes along the slinky, the coils are very close together. This set of coils that are closer together than other

compression ›

rarefaction ›

longitudinal pulse ›

coils on the slinky is called a compression, since the coils there are pushed closer together. The coils immediately to the left of the **compression** in the figure are more spread out than they normally are; this set together is called a **rarefaction**. This pulse is moving along the slinky in a direction that is parallel to the length of the slinky. It is therefore called a **longitudinal pulse**.

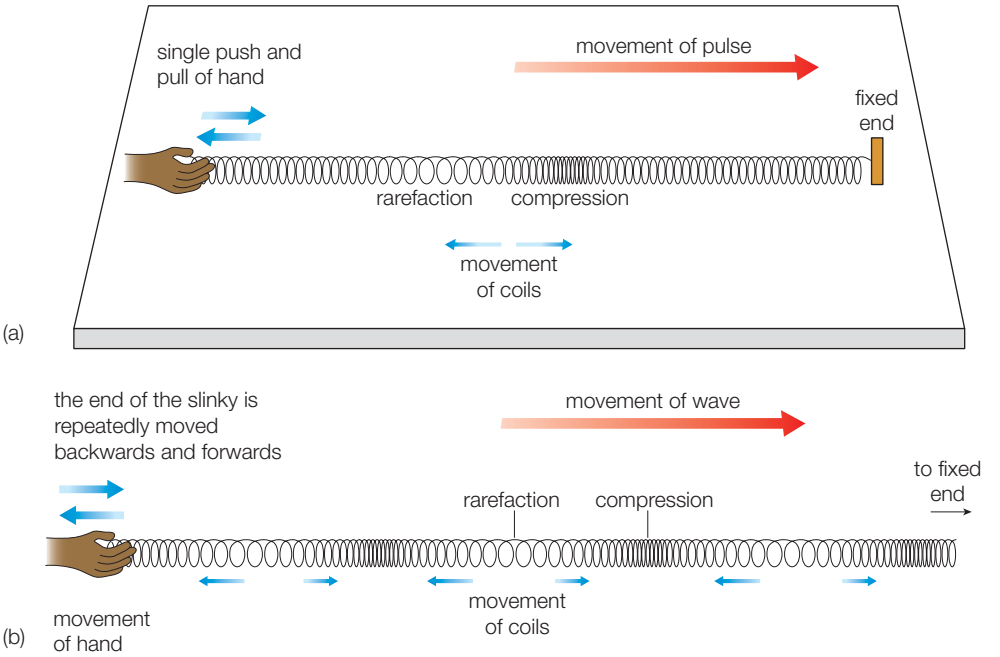


Figure 16.4 (a) A longitudinal pulse in a slinky. (b) A longitudinal wave. Note the compressions and rarefactions.

If you push the end of the slinky you are holding inwards through a distance of 30 cm, pull it back past its starting position to a point 30 cm on the other side and then return it to the starting point, just once, one cycle of a wave is produced. This pulse, too, will be seen to move along the slinky from the source to the other end. This will consist of a set of coils very close together, a compression, and, some distance away on the slinky, another set further apart than the normal separation, called a rarefaction.

The compression and rarefaction are seen to move along the length of the slinky, always the same distance apart. The pulse is a travelling or progressive pulse. As with transverse progressive waves, this pulse takes energy along the slinky away from the moving hand.

If the movement of the hand is repeated, a continuous set of pulses is seen travelling outwards from the vibrating hand, each complete oscillation of the hand giving rise to one complete wave on the slinky. A series of alternating compressions and rarefactions is seen to travel away from the source. This series of alternating compressions and rarefactions moving away from the source along the slinky is a progressive (or a travelling) **longitudinal wave**.

In contrast to transverse waves, however, the movement of each point on the slinky is not perpendicular to the direction of travel of the pulses, but parallel to it. This can be seen by tying a short length of brightly coloured string at a point on the slinky and observing its motion as the waves travel past it. The string moves to and fro in a direction parallel to the length of the slinky. This feature is what distinguishes a transverse wave from a longitudinal wave.

Since all longitudinal waves need a medium for the transmission of their energy they are therefore all matter waves. Examples of a longitudinal wave are ordinary sound and ultrasound.

IT Q2

'A vibrating body has both kinetic energy and potential energy.' What evidence can you give in support of this statement?

longitudinal wave ›

In the transverse wave the vibrations of particles in the medium is at right angles to the path of the wave; in the longitudinal wave the vibration of the particles is parallel to the path of the wave

Waves in a ripple tank

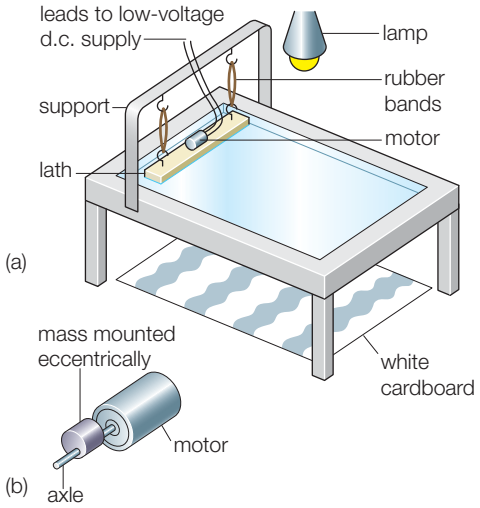


Figure 16.5 (a) The ripple tank. (b) A mass (small metal cylinder) is attached eccentrically to the axle of the motor.

We can use a ripple tank to study how water waves behave. A ripple tank is a shallow glass or plastic water tank mounted on legs. A lath is suspended by rubber bands just touching the surface of the water, as shown in figure 16.5 (a), and a motor is attached to the lath. The motor has a small metal cylinder attached eccentrically (off-centre) to its axle (figure 16.5 (b)), so that when the axle spins the motor 'wobbles'. The lath vibrates with the same rhythm (frequency) as the motor, so creating ripples in the water as it moves up and down in the surface of the water.

To make the wave effects more visible, the tank is illuminated by a lamp placed about 30 cm above the tank, the ripples being seen as shadows on a sheet of white paper or cardboard placed on the laboratory bench directly below the tank. The wave pattern seen is much larger than the actual pattern in the tank, with the crests showing up as bright areas and the troughs as darker ones. Some ripple tanks are illuminated from below, with the shadow pattern being cast onto the ceiling.

Figure 16.6 (a) shows continuous straight (plane) waves being created by a vibrating lath. The lines of crests (or troughs) are called *wavefronts* (page 257), an important concept that we shall be using in a later chapter. Figure 16.6 (b) is the view of the wavefronts as seen from above. Figure 16.6 (c) is a view of the waves as seen from the side, or a 'wave profile'.

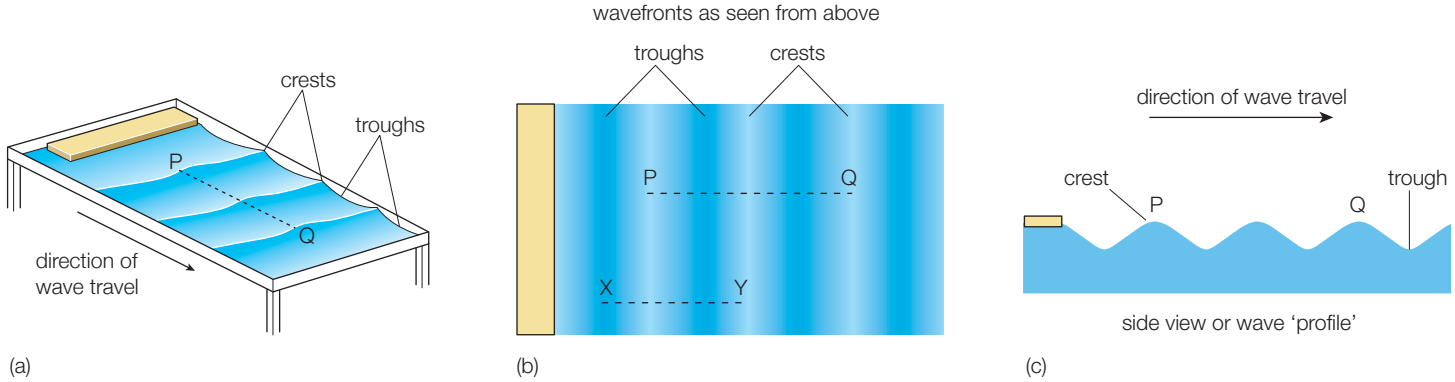


Figure 16.6 (a) Straight (plane) waves in a ripple tank. (b) View of the wavefronts from above. (c) Side view or 'profile' of the waves.

The ripples you see on the surface of a pond may actually be a combination of transverse (up and down) and longitudinal (to and fro) water waves, so the particles of water may be moving to and fro as well as up and down.

Rays

ray

The concept of 'a ray of light' is one of tremendous importance in geometrical optics to which we will be introduced in a later chapter. What is a **ray** of light or a ray of sound or a ray of any form of energy with which we may associate wave motion?

In figure 16.7, the vibrating dipper, S (something like a pencil point or a small sphere), is moving at regular intervals in and out of the water in a ripple tank and is generating waves on the water surface. The crests of the ripples appear as a set of concentric circles moving away from the source, S.

You have certainly seen something similar in pools of water into which water drops are falling on a rainy day. The arrows drawn outward from the centre (the source) represent the directions in which the transverse wave is moving, and therefore the directions along which energy is carried from the source. These arrows show the direction in which the wave energy is travelling away from the source. Each one of these lines of travel of the wave energy away from the source is called a ray. If the waves are transferring light energy

ITQ3

Sketch the wave profile between X and Y in figure 16.6 (b).

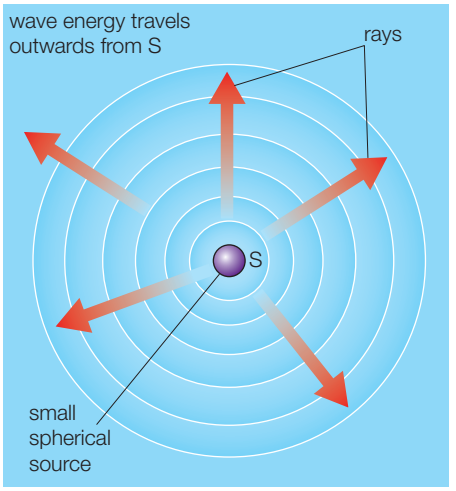


Figure 16.7 A small spherical dipper creates a series of concentric waves. The rays show the directions in which the wave is travelling, that is, the direction in which energy is being transferred.

A ray drawn from any point on a medium shows the direction of travel of wave energy at that point.

from a light source the ray is called a light ray; if the source is a source of sound, the ray is called a sound ray.

It is clear from figure 16.7 that, since the waves are circular and the rays are travelling away from the source at the centre, the rays are meeting the wavefronts at right angles. This is the case for all rays whether they are radiating from a point source (like those in the figure) or whether, as in figure 16.6, the wavefronts are parallel. In the latter case since the wavefronts (and so the waves) are parallel, and the rays are always perpendicular to them, the rays will themselves be parallel to one another. Since the rays leaving a point source like that in figure 16.7 diverge (that is, they get further and further apart as they travel away from the source), they are called diverging rays. Parallel rays are associated with a plane vibrating source and the waves associated with these parallel rays are called plane waves.

Remember that the path of the ray is always perpendicular to the wavefronts where a wavefront is any line of crests or any line of troughs found on a medium along which the wave energy is being taken.

Rays from a single source may be parallel, convergent or divergent. If they spread out from the source, the rays are divergent and the set of rays together form a divergent bundle of rays or a divergent beam. The rays in figure 16.7 are divergent. In a parallel beam, the rays are all parallel to one another; and in a convergent bundle or beam, the rays get closer together. We will see how a convergent beam of rays may be obtained later.

What happens when a wave travels through a medium?

To answer this question we are first going to consider what happens in a specific small part of the medium over time, and then take a 'snapshot' look at the entire wave at a specific instant of time.

Displacement of a given particle of the medium over time: the displacement–time graph

Earlier in this chapter we considered the motion of a small piece of string tied to one coil of a slinky. While the slinky was vibrating transversely, the movement of the string was to and fro across the length of the slinky, at right angles to the direction of travel of the wave (figure 16.8 (a)). This motion was a vibration, or oscillation. It can be represented on a graph that shows the change in position (displacement) of the string with time, as the wave passes along the slinky.

We are going to examine in detail what happens as the string, initially at its mid-position, moves up to its highest point, moves back down through the mid-position to its lowest point, and moves up again to the mid-position (figure 16.8 (b)). In other words we are going to look at a single complete oscillation. The time for one oscillation is called the **period** of the oscillation, and is often denoted by T.

period ›

Figure 16.8 (b) shows the position of the piece of string during a single oscillation, beginning with the string at the mid- or mean position (A in the diagram). We divide the time for one oscillation, T, into 8 equal parts and we next consider where the string is at intervals of $T/8$, that is $0T/8$, $1T/8$, $2T/8$, $3T/8$ and so on.

- At time zero, $0T/8$ it is still moving upwards, through B.
- At time $1T/8$ it is still moving upwards, through B.
- At $2T/8$ the string is momentarily at rest at C, its highest point.

- At 3*T*/8 the string is moving downwards through B.
- At 4*T*/8 it is moving downwards through the mean position A, where we began the observation.
- At 5*T*/8 it is moving downwards through position D.
- At 6*T*/8 the string is momentarily at rest, at its lowest point, E.
- At 7*T*/8 the string is moving back up through D.
- Finally, at 8*T*/8, it is moving upwards through A, as it was at the start of the observation.

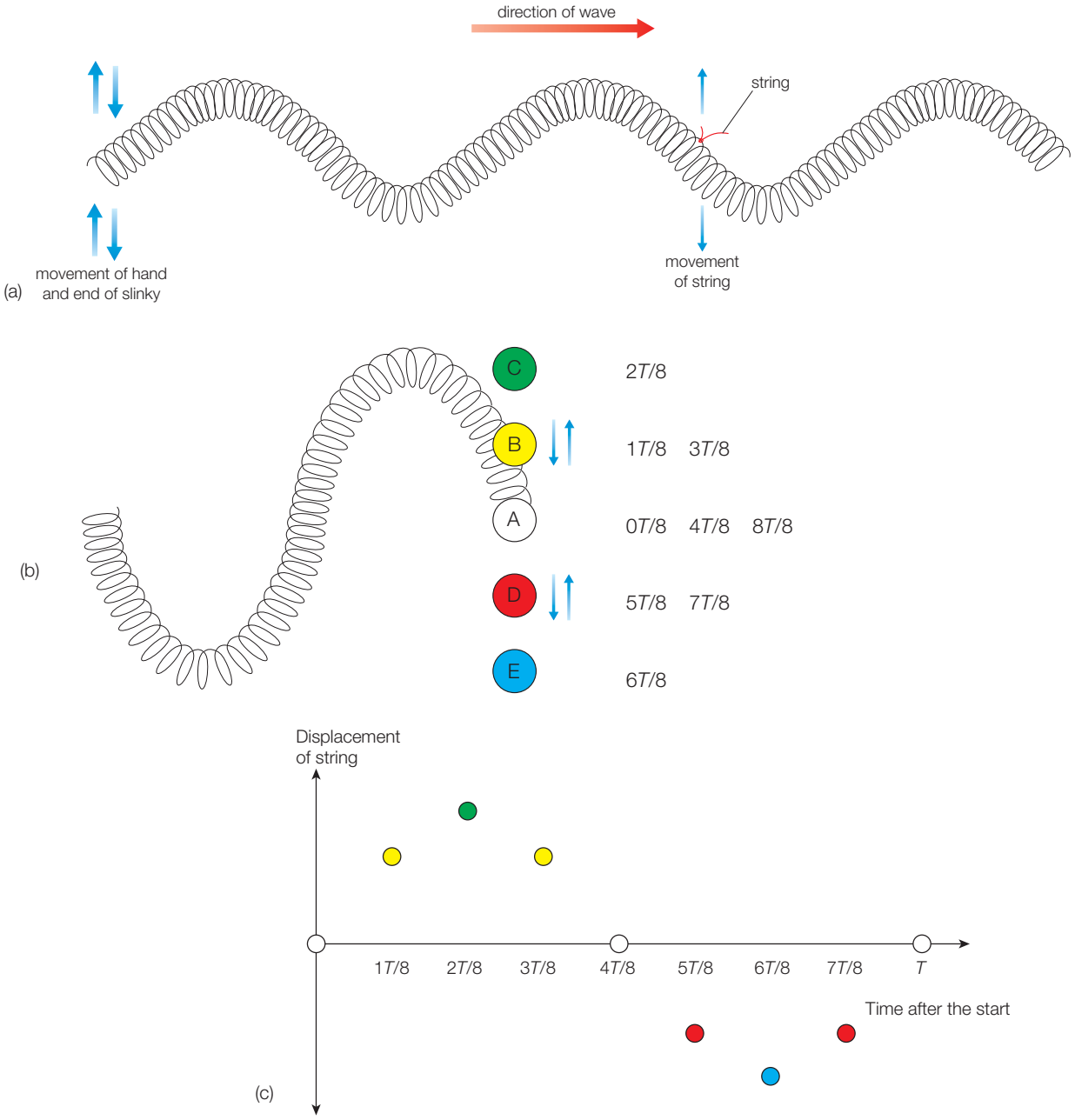


Figure 16.8 Displacement with time at a given position in a medium (slinky) as a wave passes through. The vertical scale in part (b) has been increased for clarity.

displacement–time graph ›

The graph in figure 16.8 (c) shows the position of the string at successive time intervals of $T/8$ from the start. You may notice that the graph has the same shape as the wave on the slinky. This graph is called a **displacement–time graph**, since the variable on the vertical axis represents the displacement of the string, or a single particle of the wave, from its mean position and the variable on the horizontal axis is time measured from the moment $t = 0$ when the observation began.

Figure 16.8 shows the displacement in detail for only a single oscillation, representing only one wave cycle. The graph in figure 16.9 represents the displacement over several cycles of a particle in a medium carrying a continuous wave. So the graph of figure 16.8 is really the first complete oscillation (in time T) of figure 16.9.

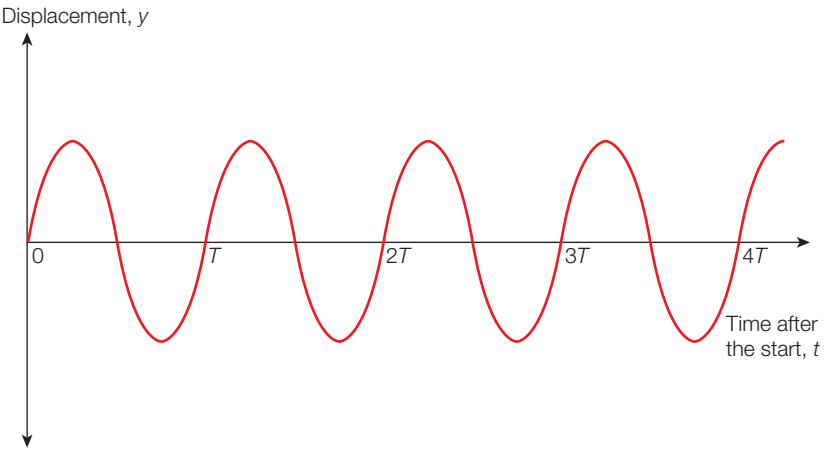


Figure 16.9 Displacement–time graph for a given particle within a medium carrying a continuous wave.

Displacement of all particles along the wave at a given moment: displacement–distance from source, or displacement–position graph

If we were to take a photograph of the wave as it travelled along the slinky at a specific moment, the picture might look like the graph of figure 16.10. This graph shows how the displacement of particles on the wave from their rest positions varies from place to place along the wave at that given instant. We see the displacement varying in a regular manner depending on how far a particle is from the origin or source of the wave.

A progressive wave varies simultaneously in both time and space. So when we represent the wave graphically, *either*

(i) the time must be held constant for all particles on the wave and x, the distance of particles from the source, allowed to vary as in the displacement–position graph ('snapshot' graph), *or*

(ii) the position (distance from the source) must be held constant, as in the displacement–time graph and the displacement shown for one particle (which means at a fixed distance, x, from the source).

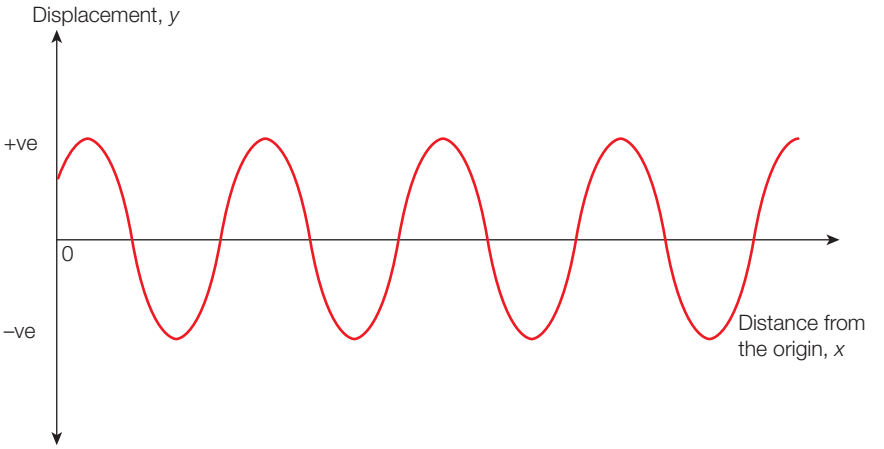


Figure 16.10 Displacement–position graph for a wave at a given instant of time.

displacement–position graph ›

The *y*-axis again shows the displacement of particles and the *x*-axis their position (distance from the source) along the wave. This is therefore a **displacement–position graph**. We shall be looking more closely at this type of graph in the next chapter.

What happens in the case of longitudinal waves?

In the discussion above, we assumed the vibrations to be transverse. The same type of graphs would be obtained if we plotted the displacement of particles on a longitudinal wave against time and against distance.

Displacement in the case of the longitudinal wave means the same thing, namely the distance of a particle on the wave from its mean position or the position it occupied before the wave started. It does not matter whether the displacement is parallel to the direction of travel of the wave or perpendicular to it. The way in which the displacement changes with time is exactly the same as for transverse waves. For longitudinal waves, as with transverse waves, the displacement to one side of the mean position is positive and that to the other side is negative for all the particles on the wave (see figure 16.11).

Displacement–time and displacement–position graphs look the same for both transverse and longitudinal waves. For longitudinal waves the graphs represent displacement parallel to the direction of the wave; for transverse waves the displacements represented are perpendicular to the direction of the wave.

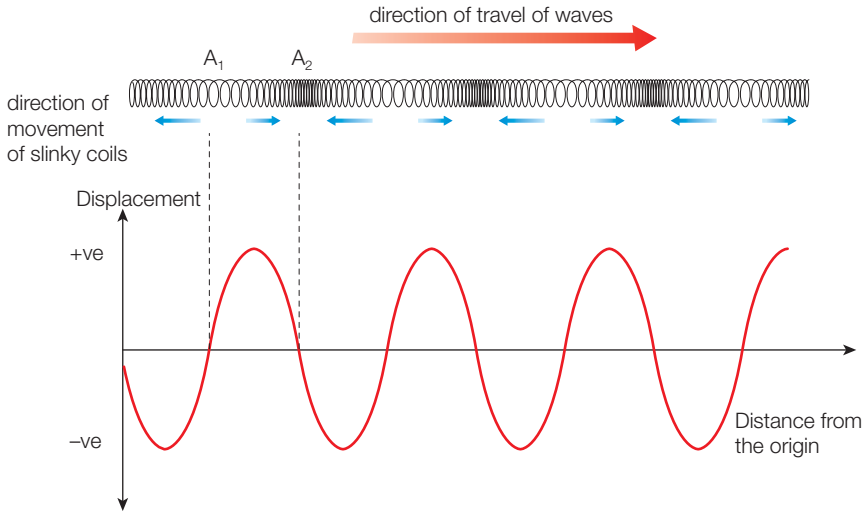


Figure 16.11 Displacement–position graph for particles in a longitudinal wave in a slinky. The displacement is parallel to the direction of travel of the wave. At the points of greatest rarefaction or compression, such as A_1 and A_2, the coils of the slinky are moving through their mean position of zero displacement. Displacements to the right are positive and those to the left are negative.

Figure 16.11 shows a displacement–position ('snapshot') graph for a longitudinal wave, for example in a slinky. A displacement–time graph for a given particle in a longitudinal wave would be exactly the same as the displacement time graph for a transverse wave, as shown in figure 16.9.

Matter waves and electromagnetic waves

Matter waves

Matter is substance, that is, anything that has mass.

When a wave is transmitted by matter, small 'bits' of that matter vibrate either at right angles to the path of the wave (if the wave is transverse) or parallel to the direction of travel (if the wave is longitudinal). When particles

of the material vibrate, they possess kinetic energy; and when they slow down and come to rest momentarily at the extreme points of their motion, all the kinetic energy they had while moving is converted into potential energy. Thus the energy of the vibrating particles of the medium will be potential, kinetic or a mixture of the two. Since these energies are due to mechanical forces (pushes and pulls, like tension) in the medium, we say that the energy associated with their transmission is mechanical in nature (see chapter 6). So matter waves or material waves transmit mechanical energy. Examples are waves travelling on a rope, on a slinky or on a water surface and in the air.

CHAPTER 6

Electromagnetic waves

In the case of the waves that reach us from the Sun, however, most of their journey to Earth is made through space where there is no matter at all. So if all waves take energy with them as they travel, what sort of energy is transported by these waves? It is certainly not mechanical! The answer lies in the way the waves are produced in the first place. We will leave further discussion of the nature of these waves for chapter 19.

CHAPTER 19

All that we will say about them for the time being is that

- they can all travel in a vacuum;
- they all travel at the same speed in a vacuum – this speed is $3.0 \times 10^8\,\mathrm{m\,s^{-1}}$;
- they are all transverse.

Chapter summary

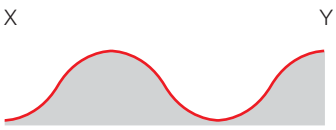
- A vibration (or an oscillation) is a to-and-fro motion of a body about a mean position.
- A vibrating body has kinetic energy and potential energy.
- A pulse is produced in a medium when an oscillation or part of an oscillation performed by a source transfers energy from the source through the medium.
- A wave is produced when vibrations from the source transfer energy continuously through the medium.
- A matter wave is produced when an oscillation passes its energy continuously into a material medium that is in contact with it.
- A wave is a disturbance that repeats itself at all points along the path that it follows.
- There is continuous oscillation at every point along the path of a wave.
- There are two kinds of waves, transverse and longitudinal.
- In a transverse wave the vibrations take place in a direction perpendicular to the direction of travel of the wave.
- In a longitudinal wave the vibrations take place in a direction that is parallel to the direction of travel of the wave.
- Electromagnetic waves consist of transverse vibrations.
- Longitudinal waves travel only in material media, and so all longitudinal waves are matter waves.
- Material media can also transmit transverse waves.
- The direction along which wave energy travel is called a ray.
- A progressive transverse (or travelling) wave is one in which the ‘high’ points (called crests) and the ‘low’ points (called troughs) seem to move away from the source of the wave.
- A progressive longitudinal wave is one in which points of high particle density (called compressions) and points of low density (called rarefactions) seem to move away from the source of the wave.

Answers to ITQs

ITQ1 The energy is taken by the single pulse along the slinky.

ITQ2 Both kinds, potential and kinetic, since the slowing down of the vibrating source as it approaches one extreme position will mean a loss of kinetic energy. This loss of kinetic energy will, by the conservation law, be accompanied by a gain of potential. Thus for most of the time the energy is both kinetic and potential. So the energy changes of the body are PE ➞ KE ➞ PE.

ITQ3



Examination-style questions

1 I stand in one place and move my right hand up and down at my side in a vertical plane. I then begin to walk forwards slowly, then faster and then faster still with my hand still moving up and down at the same rate. Sketch a diagram showing the shape of the figure my hand might trace out from when I begin walking.

2 I stand at the shallow end of a swimming-pool facing the near edge of the pool. With the palm of my hand in a vertical plane at the surface of the water, parallel to the near edge, I repeatedly move my palm towards and away from the near edge of the pool. My palm is in a vertical plane throughout these oscillations. What type of waves am I likely to produce on the surface of the water? Give a reason for your answer.

3 Imagine you have accidentally kicked your football into a nearby 'trench'. You try to recover the ball by throwing a few fair-sized rocks into the trench behind the ball in the hope that the ball will move to your side of the trench, but it does not. Describe what happens to the ball when a rock is thrown into the trench and explain why the ball does not move towards you.

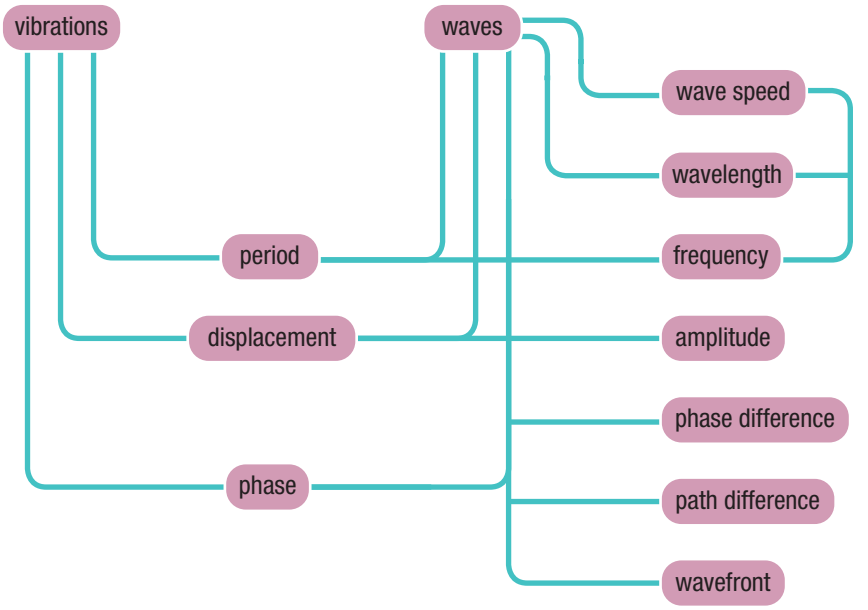
4 State the reason or reasons why you would or would not agree that a coconut tree oscillates when a strong steady breeze is blowing.

5 You accidentally knock an empty bottle standing on a table but not hard enough to make it fall over and you notice that the bottle carries out a motion that you think you can describe as an oscillation before it finally comes back to rest. State the reasons you would give for arguing that the bottle oscillated before it came to rest.

17 Wave Characteristics

By the end of this chapter you should be able to:

- define and use the following terms: frequency, period, wavelength, displacement, amplitude, phase, phase difference, path difference, wavefront, wave speed
- explain the meanings of and use the expressions 'in phase' and 'out of phase'
- recall the relationship between the amplitude and the energy of a wave
- recall and use the relationship $v = f\lambda$, where v is the wave speed, f is the frequency, and λ is the wavelength



Introduction

When talking about cricket we use terms like 'cover point', 'square leg' or 'bowler', and a listener who is familiar with the sport knows exactly what we mean. Similarly, in athletics we may use words like 'sprint', 'javelin' or 'marathon'; each sport has its own vocabulary. In the same way, when we describe waves and their behaviour, we use specific terms such as 'frequency' and 'wavelength'. These are called **wave parameters** or **wave characteristics**. In this chapter we will discuss the significance of all the terms mentioned in the concept map, including phase, phase difference and path difference, since the effects we call interference and diffraction depend upon a firm understanding of these concepts. We begin with the most basic terms, frequency and period.

wave parameters ›
wave characteristics ›

Frequency and period

Frequency

We have seen in the previous chapter that waves are caused by some kind of oscillation or vibration. We know, too, that for every complete vibration, one complete cycle of disturbance (see figure 17.1) is produced in the medium if the wave is a material wave. Thus if the source makes two vibrations each second, then two waves will appear in the medium every second. The number of vibrations made by the source and, therefore, by each particle of the medium per second is called the **frequency** of the wave. The symbol for frequency is f.

The frequency of a wave is either: (i) the number of oscillations carried out each second by a particle of the transmitting medium, or (ii) the number of waves that go past a given point in the medium in 1 second. Note that the frequency of the particles on the medium is exactly the same as that of the body that oscillates to produce the waves, the oscillator.

frequency ›

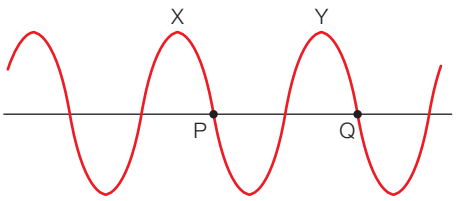


Figure 17.1 The portion of the wave between the particles at X and Y represents one complete cycle of the wave, since: (i) the particles at X and Y are in similar situations, both being at a crest, and (ii) X and Y are consecutive. The portion between P and Q is also one cycle, as P and Q are also consecutive points where the particles are in similar situations, that is both moving in the same direction through their equilibrium (zero displacement) position.

IT Q1

What is another name for a 'travelling' wave?

If the wave is a travelling transverse wave, then the crests and troughs on the wave will be seen to move along the medium away from the source. We can also think of the frequency of a wave as the number of wave cycles that go past a given point in the medium in 1 second, since the waves move away from the source as fast as they are formed at the source.

Generally, we say that:

> **The frequency, f, of an oscillator is the number of oscillations (or cycles) it performs in one second.**

cycles per second ›

hertz ›

Since frequency is defined as number of **cycles per second**, its S.I. unit will be: '/second' or 'second^{-1}'. Using symbols, that is '/s' or 's^{-1}'. The unit 'per second' or 's^{-1}' has been given the name '**hertz**' (abbreviation Hz) after Heinrich Hertz, a German physicist who, in the late 19th century, was the first to produce electromagnetic waves experimentally (see chapter 19) .

The abbreviation for hertz is Hz and not H because another unit called the henry, used in electricity, has H as its symbol.

We also use multiple units of the hertz, such as the kilohertz (kHz, 10^3 Hz), the megahertz (MHz, 10^6 Hz) and the gigahertz (GHz, 10^9 Hz), the last of these units being very common in computer technology. The speed of your computer's microprocessors is often measured in gigahertz.

IT Q2

A lath moving up and down in a ripple tank takes 0.25 s to make one complete oscillation. What is the frequency of the waves it produces?

Period

Consider a wave with frequency f travelling in a medium. This means that points on the wave will vibrate f times in 1 second. Thus the time for one oscillation is $1/f$ of a second. This is called the period of the wave (page 246), and is denoted by the letter T.

MATHEMATICS: 'finding for one'

> **The period of a wave, T, is the time taken for one full oscillation to be completed:**
> $T = \frac{1}{f}$

The unit for period is the S.I. unit of time, the second.

These definitions apply to all waves, whether transverse or longitudinal, since oscillations take place in both types of wave.

ITQ3

Imagine that a picture taken of a longitudinal wave showed compressions and rarefactions as in figure 17.2. C represents a compression and R a rarefaction.
The frequency of the wave is 1000 Hz.
(i) How many complete cycles are there between C_1 and R_3?
(ii) Calculate how long it takes:
(a) compression C_1 to reach the position occupied by compression C_3;
(b) rarefaction R_1 to reach the position occupied by rarefaction R_3; and
(c) compression C_2 to reach the position occupied by rarefaction R_3.

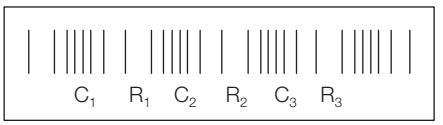


Figure 17.2

In a transverse progressive wave, we could think of the period as the time for a crest at a certain point on the wave to be replaced by the next crest at the same point. In the case of a longitudinal wave, it would be the time taken for a compression at any point on the wave to be replaced by the next compression at that point.

Worked example 17.1

A sound wave is an example of a longitudinal wave. A whistle emits a note of frequency 600 Hz.

(i) How many compressions arrive at the eardrum of a listener who listens to the note for 15 s?

(ii) What is the time interval between the arrival of successive compressions?

Solution

If the frequency of the note is 600 Hz, then 600 complete waves are transmitted in the air between the whistle and the listener each second. There will therefore be 600 compressions and 600 rarefactions leaving the whistle each second and the same number arriving at the listener's eardrum in that time.

(i) In 15 s, 600 × 15 compressions will arrive at the listener's eardrum, i.e. 9000 compressions.

(ii) The time interval between the arrival of the compressions is the time it takes to produce one wave, or the period, T, which is

$$T = \frac{1}{f} = \frac{1}{600}$$
$$= 0.0017\text{ s (2 sig. fig.)}$$
$$= 1.7\text{ ms}$$

Phase and wavelength

Phase

This is an important concept.

Figure 17.3 (a) shows a wave at a particular moment. The section between P_1 and P_2 represents one complete cycle of the wave and so does the section between P_2 and P_3. The particles at P_1, P_2 and P_3 are at successive crests. They are at the same stage in their vibration, and we say that they are '**in phase**' with each other. The particles at the troughs, S_1, S_2 and S_3, are also in phase with each another. We can also see that the sections of the wave between Q_1 and Q_2, and also between Q_2 and Q_3, represent one complete cycle, so the particles at Q_1, Q_2 and Q_3 are at the same stage in their vibrations and are in phase with each other.

in phase ›

We could say then that:

Two or more points on a wave are in phase if the disturbance on the wave is at the same stage at those points.

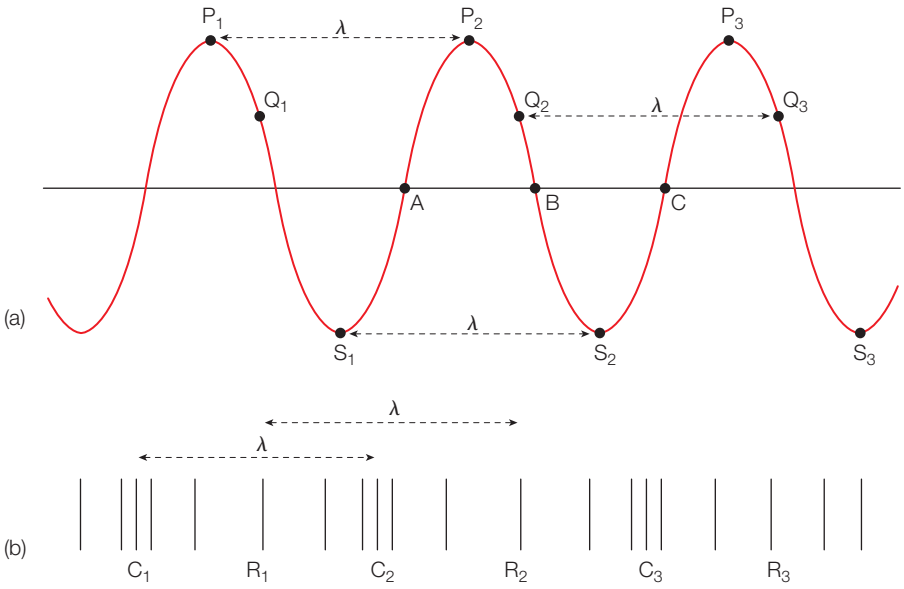


Figure 17.3 (a) Particles at the same stage of their vibration, like P_1, P_2 and P_3, are said to be 'in phase'. (b) In longitudinal waves also, particles at the same stage in their oscillation, such as those at the centres of compressions like C_1, C_2 and C_3, are all in phase with one another.

IT Q4

The dotted blue graph in figure 17.4 represents the profile of a transverse wave at a certain moment. The arrows show the directions in which particles are moving in order to give the wave profile shown in the red graph shortly afterwards.

(i) Group together the sets of points that are in phase with one another.

(ii) Give four pairs of points that are out of phase with each other.

(iii) Which pairs of points are directly out of phase (in antiphase)?

out of phase ›

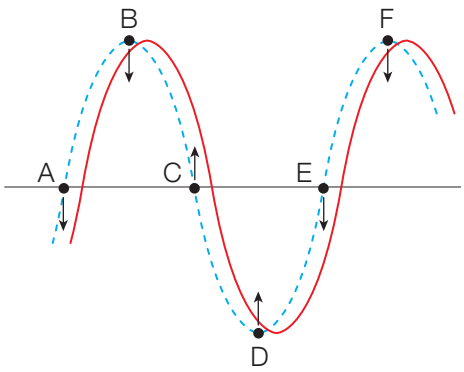


Figure 17.4

The particles at, for instance, P_1 and S_1, or at P_2 and Q_2, are at different stages of their oscillation and we say that they are '**out of phase**'.

The particles at A, B and C in figure 17.3 (a) are all at their mid- or equilibrium position. However, looking carefully, we can see that the condition of particles at A and C are similar, but not that of particles at A and B. Because of the similarity at A and C, the particles of the medium in these positions will be in phase, but the condition of particles at A and B will not. Whereas the graph at A and at C has a negative slope, at B the slope is positive. This could be interpreted to imply that at A and at C the displacement is changing negatively with time, but at B is behaving positively. This could be interpreted to mean that the change taking place at A and C is exactly opposite to that taking place at B.

This can be verified by sketching the profile of the wave shortly after the profile shown was obtained. It you do this you will find that in the new profile A and C are lower than before, while B is higher. This clearly shows that A and B were at the same stage (doing the same thing), but B was at an exactly opposite stage (doing the exact opposite). This shows beyond doubt that A and B are indeed in phase, but B is exactly out of phase with both. While the particles at A and C are moving downwards, that at B is moving upwards. Since the particles are behaving in exactly opposite ways, we could say that they are exactly opposite in phase.

MATHEMATICS: algebra – slope of a graph

If we take a point at a crest, like P_1, and one at a trough, like S_2, we find that they are at opposite stages of their motion. The same sketch profile taken moments later would show this since. the particle at P_1 is about to move downward while that at S_2 is about to move upward. We say that their phases are opposite or that they are '**directly out of phase**'. When particles are directly out of phase with each other, they are sometimes said to be '**in antiphase**'.

directly out of phase ›
in antiphase ›

These terms may also be applied to longitudinal waves (figure 17.3 (b)). Particles at the centres of compressions C_1 and C_2 are in phase; they are at the same stage of their oscillation. The particles at R_1 and R_2, which are centres of rarefactions, are also in phase with each other. Particles at C_1 and C_2 are, however, in anti-phase with particles at R_1 and R_2.

Wavelength

wavelength ›

Looking again at figure 17.3 (a), we notice that the distance between crests P_1 and P_2 is the same as that between P_2 and P_3. This distance is called the **wavelength** of the wave and is denoted by the symbol λ ('lambda'). The distance between consecutive troughs is also one wavelength.

The symbol for wavelength, λ, is a Greek letter, spelt 'lambda', and pronounced 'lam-da'.

The wavelength, λ, is the distance between two consecutive points on a wave which are exactly in phase.

Note the word 'consecutive'. The points chosen must occur one after the other and be in phase.

The wavelength of a wave is therefore the length of one cycle of that wave. It could be the distance from one crest to the next crest of a transverse wave or the distance from one compression to the next on a longitudinal wave, as in figure 17.3 (b).

Since wavelength is a distance, its S.I. unit is the metre. Any convenient sub-unit (mm, cm) may be used if this is more suitable.

ITQ5

Figure 17.5 shows a wave profile for a part of a progressive wave on a rope.

(i) Group together the points which are: (a) in phase; (b) directly out of phase.

(ii) Which distances represent: (a) one wavelength; (b) two wavelengths; (c) just less than half a wavelength; (d) a little more than half a wavelength?

(iii) The number of cycles shown is approximately how many?

A 2 B $2\frac{1}{4}$ C $2\frac{1}{2}$ D 4

amplitude ›

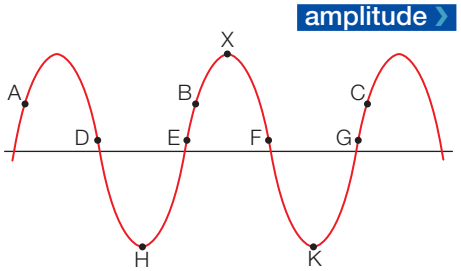


Figure 17.5

Displacement and amplitude

In chapter 16 we considered how the displacement of particles in both transverse and longitudinal waves varied with time and with distance from the source (pages 246–248). The maximum displacement of a particle on a wave is called the **amplitude** of the wave (figure 17.6). The amplitude of a wave is an important parameter – important because it determines the energy in the wave. The greater the amplitude of the wave, the greater is the energy transmitted by it. It is important that you remember this as we shall see in the next chapter.

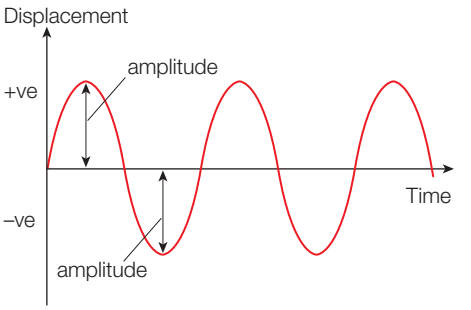


Figure 17.6 The maximum displacement of a particle on a wave is called the amplitude of the wave.

Phase difference and path difference

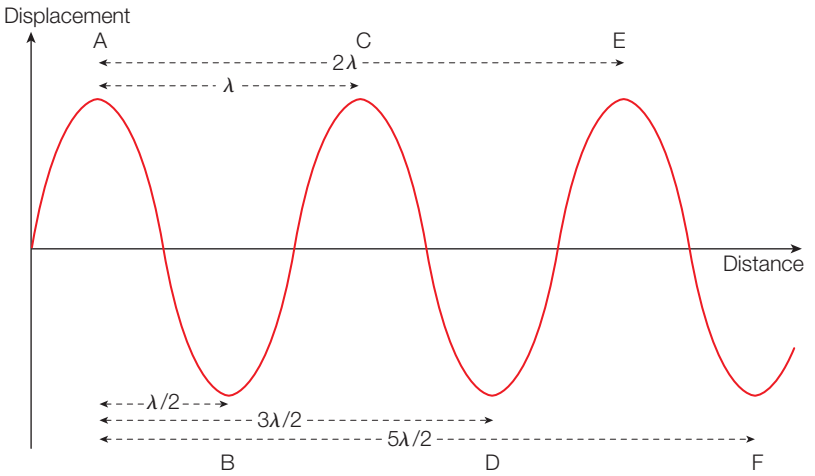


Figure 17.7 Points separated by a whole number of wavelengths, like A and C, are in phase. Points separated by an odd number of half-wavelengths (like A and B or A and D) are directly out of phase.

We have seen that two particles on a wave which are in the same situation at a particular moment, or at the same stage of their vibrations, are said to be in phase. Consecutive particles that are in phase are separated along the wave by a distance of one wavelength (1λ). Particles separated by distances of 2λ, 3λ and so on will also be in phase with one another (figure 17.7). In general, particles on a wave which are in phase are separated by a whole number of wavelengths. We write a whole number of wavelengths as $n\lambda$, where n is a whole number or an integer.

From figure 17.7, you can see that particles that are directly out of phase ('in antiphase') are separated by λ/2, 3λ/2, 5λ/2, and so on. In other words, particles that are directly out of phase are separated by an odd number of half-wavelengths.

phase difference ›

path difference ›

So for the wave shown in figure 17.7, the **phase differences** between the particle at A and the particles at B, C, D, E and F depend on how much further the wave has travelled past A to reach these points. Thus the path to B is λ/2 greater than the path to A, and we say that there is a **path difference** between A and B of λ/2. This path difference of one half-wavelength means that A and B are directly out of phase. We can therefore associate phase differences with corresponding path differences. A path difference of an odd number of half-wavelengths corresponds to vibrations being directly out of phase. We will need to be clear about the significance of path difference and also phase difference when we come to consider the important topic called interference.

ITQ6

A lath makes regular vibrations in a shallow trough, thereby generating plane waves of wavelength 4 cm on the surface of the water (figure 17.8). A, B, C, D and E are points on the water surface 4 cm, 7 cm, 12 cm, 18 cm and 19 cm from the vibrating lath. Fill in the blanks in the table, using the example of the first row.

Particle	In phase with A	Out of phase with A	Path difference relative to A
B		✓	3/4 wavelength (or 3λ/4)
C			
D			
E			

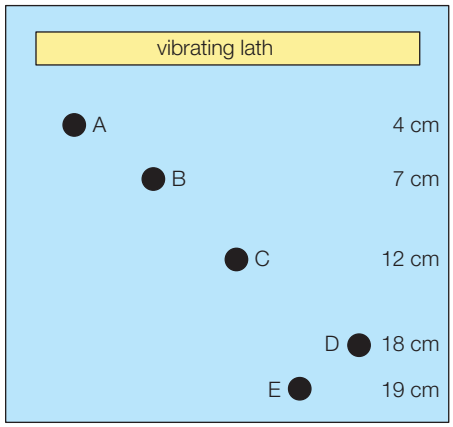


Figure 17.8

Wavefronts

We met this term in the last chapter and commented on its importance in understanding the theory of waves.

The waves produced by a lath vibrating with a constant frequency on the surface of water in a ripple tank travel away from the lath. Lines of crests can be seen moving away from the vibrating lath, and these lines of crests are evenly spaced if the trough is of uniform depth. The particles of water along the top of a particular line of crests will all be the same distance from the source, since all of the crests were created at the same time. The same thing can be said of the particles along the bottom of any given line of troughs.

Wavefront

A line of crests (or a line of troughs) which were all created at the same time at the source is known as a **wavefront**. The wavefront is the shape of the wave as seen from above (see pages 245 and 246).

Wavefronts are always 'on the move', since crests and troughs, and compressions and rarefactions, are constantly moving away from the source that created them.

A wavefront need not be a line of crests or even a line of troughs. It is any line of particles whose motion was created at the same time at the source of the wave. Such particles will naturally all be in phase. They are therefore all at the same stage of their motion, as well as being the same distance from the source (if the ripple tank is of uniform depth).

The particles on a wavefront are all the same distance from the source of the waves in any one medium.

If we are considering a series of wavefronts, the points on all the wavefronts must be at the same stage of their movement. This means, for example, that we can consider either lines of crests or lines of troughs, but we may not consider some lines of crests together with some lines of troughs.

Shape of wavefronts

For material waves, the shape of wavefronts will depend on the shape of the source. If the waves are produced on the surface of a body of water and the source is the point of a vibrating pencil, the lines of crests or wavefronts will move away as concentric circles. The distance between the wavefronts will, clearly, be equal to the wavelength of the waves. If the pencil is vibrating in a regular fashion, the distance between consecutive wavefronts will be constant.

If, however, the source is a straight lath vibrating regularly, then the wavefronts will be parallel straight lines with a constant distance between consecutive wavefronts. Again, the distance apart of the wavefronts will represent the wavelength of the waves.

Wavefronts are often referred to simply as 'waves'.

A source of sound waves like a whistle, with the medium surrounding it, will produce waves which will leave in all directions about the source and the wavefronts will be concentric spheres, the common centre of these spheres being the source of the sound.

Describing waves and wavefronts

The shape of the wavefronts is used to describe the shape of the wave. For example, if a source produces waves with circular wavefronts, we describe them as **circular waves**. **Plane waves** are those whose wavefronts are straight lines.

plane wave circular wave ›

Rays and wavefronts

The relationship between rays and wavefronts was discussed in the last chapter.

CHAPTER 16

If a ray is the direction along which wave energy travels (see page 245), then it is perpendicular to the wavefronts at all points along its path. We will need to remember this when we consider reflection and refraction of waves later in this section. The distance that separates consecutive wavefronts is the wavelength.

Wave speed

Suppose a vibrating source is producing f waves every second on the surface of a shallow ripple tank (figure 17.9). Suppose too that every crest produced moves away from the source with a speed v. This speed, v, is called the **wave speed**.

wave speed ›

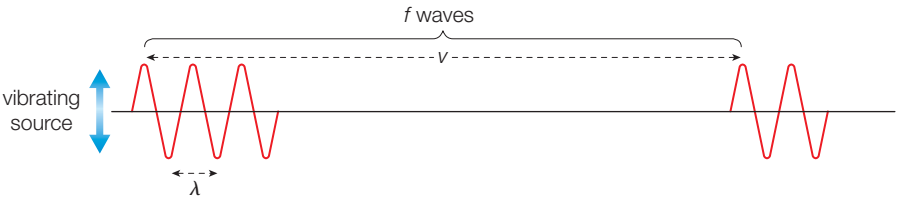


Figure 17.9 The source is producing f waves per second, and each crest moves away from the source with speed v.

So, one second after the source begins to generate a wave, the disturbance has reached a point on the water distant v from the source. If there are f waves occupying the distance v, and each of these waves is of length λ, then the distance occupied by the f waves that leave the source in one second (which is the definition of speed), v is given by

$v = f\lambda$

This equation says that:

wave speed = frequency × wavelength

This is a very important equation which applies to both transverse and longitudinal waves. It will be used very often later in this section. The S.I. unit of speed is the metre per second ($m\,s^{-1}$).

Worked example 17.2

Assuming that sound travels at $300\,m\,s^{-1}$, calculate the wavelength of a note of frequency 500 Hz.

Solution

wave speed = frequency × wavelength

Rearranging the equation:

$$\text{wavelength} = \frac{\text{(wave speed)}}{\text{(frequency)}}$$

or $$\lambda = \frac{v}{f}$$

Substituting, $$\lambda = \frac{300\,m\,s^{-1}}{500\,Hz}$$

$$= 0.6\,m \text{ or } 60\,cm$$

The wavelength of the note = 60 cm.

Chapter summary

- Wave parameters or wave characteristics are the quantities we use to describe a wave or its behaviour.
- The frequency of a progressive (or travelling) wave is either:
  - the number of complete vibrations made by a particle on that wave in 1 second, or
  - the number of cycles of the wave that go past a point on the path of the wave in 1 second.
- The period of a travelling wave is:
  - the time taken for a vibrating particle on the wave to carry out one complete oscillation, or
  - the time taken for a compression (for a longitudinal wave) or a crest (for a transverse wave) to be replaced by the next compression or the next crest, at a given point on the path of the wave.
- The frequency, f, of an oscillation or a wave is equal to the reciprocal (or the inverse) of the period, T, of the oscillation or the wave, In other words, $f = 1/T$ and $T = 1/f$.
- The unit of frequency is the hertz, abbreviation 'Hz'.
- The phase of a particle in a wave is the stage of vibration of the particle at any given moment.
- Particles in a wave that are at the same stage of their vibration at any given moment, i.e. in step with one another, are said to be in phase.
- Particles that are at different stages of their oscillation at any given moment are said to be out of phase.

- The wavelength is the distance between two consecutive points on the wave that are in phase.
- The displacement of a particle on a wave is the distance of that particle away from its normal rest position at any moment during the passage of the wave.
- The maximum displacement of any particle on a wave is called the amplitude of the wave.
- The amplitude of a wave determines the amount of energy the wave possesses.
- Two particles on a wave that are at different stages of their vibration phases are said to have a phase difference.
- The phase difference between two particles on a wave is related to the distance that separates the two particles. This separation is called a path difference.
- If two particles on a wave are exactly in phase, their path difference is a whole number of wavelengths (or an even number of half-wavelengths).
- If two particles on a wave are directly out of phase (in antiphase), their path difference is an odd number of half-wavelengths.
- A wavefront of a progressive wave is any line or surface joining disturbances on the wave that were created at the same time at the source. Such disturbances will be in the same phase.
- The speed of a travelling wave, called the wave speed or the wave velocity, is the speed at which a crest (for a transverse wave) or a compression (for a longitudinal wave) moves away from the source.
- The speed of a wave is given by the equation:
  wave speed = frequency × wavelength
  or, $v = f\lambda$.

Answers to ITQs

ITQ1 A progressive wave

ITQ2 4 Hz

ITQ3 **(i)** Two complete wave cycles **(ii)** **(a)** 2.0 ms, **(b)** 2.0 ms, **(c)** 1.5 ms

ITQ4 **(i)** (A,E), (B,F)

(ii) Any four of (A,B), (A,C), (A,D), (A,F), (B,C), (B,D), (B,E), (C,D), (C,E), (C,F), (D,E), (D,F), (E,F)

(iii) (A,C), (B,D), (C,E), (D,F)

ITQ5 **(i)** **(a)** (A,B,C), (D,F), (E,G), (H,K); **(b)** (X,H), (X,K)

(ii) **(a)** AB, BC, HK, DF, EG; **(b)** AC; **(c)** EF; **(d)** DE, FG

(iii) C $2\frac{1}{2}$

ITQ6

Particle	In phase with A	Out of phase with A	Path difference relative to A
B		✓	¾ wavelength (or 3λ/4)
C	✓		2 wavelengths (or 2λ)
D		✓	3½ wavelengths (or 7λ/2)
E		✓	3¾ wavelengths (or 15λ/4)

Examination-style questions

1 A vibrator S on the surface of a shallow trough of water makes 5 oscillations every second, generating travelling waves. The two markers, P and Q, above the trough along the line of travel of the waves, are 20 cm apart. The speed of the waves on the water is 10 cm s^{-1}. Calculate:

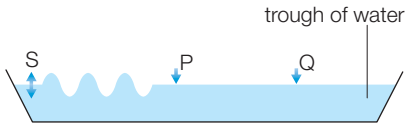
(i) the period of the waves;

(ii) the wavelength of the waves;

(iii) the number of complete waves between the points P and Q at any moment;

(iv) the phase difference between points on the wave directly below P and Q;

(v) the time taken for a crest at P to travel to Q.



2 The diagram below shows part of a progressive wave on a rope.

(i) Bracket together the particles on the rope which are:

(a) in phase;

(b) directly out of phase.

(ii) Write down, in terms of the wavelength, λ, of the wave, the values of the path difference between particles at:

(a) A and B

(b) A and D

(c) B and D

(d) C and F

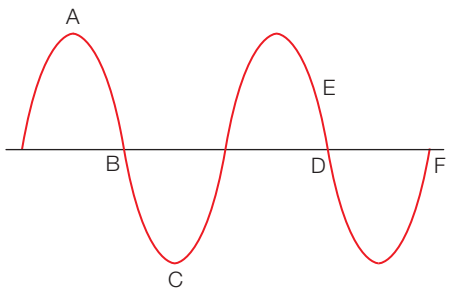
(iii) The frequency of the wave is 2 Hz. How long does it take for the crest at A to move to:

(a) a position above B;

(b) a position above C;

(c) a position above F.

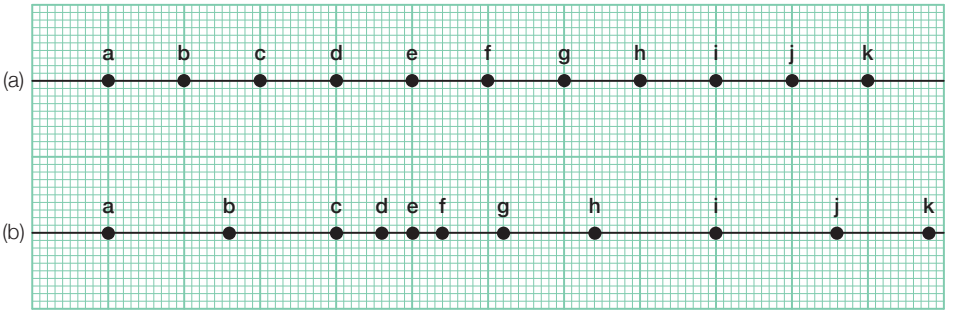
(iv) Estimate the path difference between particles at E and F.



3 A pencil point makes 5 vibrations per second in the surface of a trough of water. The point vibrates for 2 seconds and then stops. The waves travel at a speed of 10 m s^{-1}. Draw to scale the pattern of wave crests on the water 3 seconds after the oscillations start. Show all important distances on your diagram. Label:

- the first wave generated by the source;
- the last wave generated; and
- the sixth wave generated.

4 The points a, b, c, … i, j, k in diagram (a) represent 11 equidistant and highest point on the coils of a slinky which has been stretched and is resting on a smooth horizontal surface. Diagram (b) shows these points some time after a wave begins to move along in the slinky.



(i) What type of wave is travelling on the slinky?

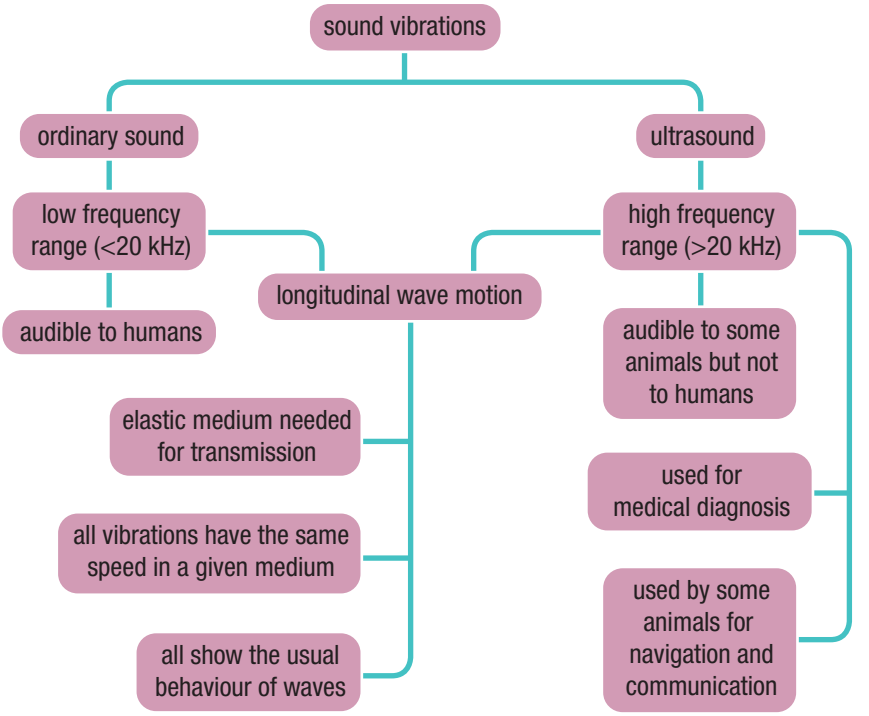
(ii) Plot a graph of displacement against position for the particles a to k for the moment represented by diagram (b). Take displacements to the right as positive. If you identify the wave as a longitudinal one, mark in the centres of compression and rarefaction.

(iii) Measure and state the values of the amplitude and wavelength of the wave.

18 Sound as Wave Motion

By the end of this chapter you should be able to:

- recall that sound is a longitudinal vibration and therefore requires a medium for propagation
- recall that sound may be produced by vibrating systems
- explain why a medium transmitting sound waves must be elastic
- describe how sound is propagated in a medium
- recall that sound waves show the effects of reflection, refraction, diffraction and interference
- cite evidence that sound waves undergo reflection and refraction and show diffraction and interference effects
- describe an experiment to show that sound is not transmitted through a vacuum
- describe a simple experiment to estimate the speed of sound in air
- recall the order of size of the speed of sound in air and apply this knowledge to practical situations
- recall that the speed of sound in a given medium depends on physical properties of the medium such as stiffness and density
- use the terms 'pitch' and 'loudness' and relate them to wave parameters
- recall the range of frequencies normally detectable by the human ear
- recall that ultrasound has a higher range of frequencies than ordinary sound, but is inaudible to the human ear
- recall the uses of ultrasound



Introduction

When someone says 'I hear a sound', we naturally think of the sensation of sound, experienced via the ear, the auditory nerve and the brain. We scarcely ever think about the mechanism that brought about this sensation of hearing, that is, the waves carrying sound energy from the source of the sound to our ears. This chapter will deal with the generation of sound waves, their transmission through a medium and the effects they show. We will see that there are two 'types' of sound, one which we may call **ordinary sound** with which we are very familiar, since we hear it every day and which we may regard as commonplace, and the other which is certainly not commonplace, but is only found in medical laboratories and in hospitals, called **ultrasound**. This form of sound has a much higher range of frequencies than ordinary sound. We will discover, too, that, like other types of waves, ordinary sound can be reflected, refracted, diffracted and made to interfere, under suitable conditions.

ordinary sound ›

ultrasound ›

Production and propagation of sound waves

Sound may be produced by having a body vibrating in a medium that is sufficiently elastic (springy) to carry the resulting waves along the medium. For example, if a wire is stretched between two fixed supports and plucked, it will vibrate. As the string moves away from its rest position to position A (see figure 18.1), it pushes the molecules of air in front of it, producing a compression. The air in the **compression** now has a higher density. The wire then moves to its other extreme position, B, and the air previously compressed now expands beyond its original volume (remember that the air is 'springy'), thereby reducing its density to a value below normal. There is now a **rarefaction** where there was previously a compression. The vibrating wire returns to its mid-position and the whole cycle is repeated. Thus the continuous vibration of the source of the sound, the vibrating wire, will produce in the air a continuous series of compressions and rarefactions. We call this series of compressions and rarefactions a **train of sound waves**. Like all waves, it will take energy from the wire, the source, to a possible receiver, like someone's ear.

compression ›

rarefaction ›

train of sound waves ›

A higher density of air molecules means that the air pressure is greater, and similarly a lower density means that the air pressure is less. Thus we can think of the sound wave as a pressure wave travelling through the air. At the centre of a compression the pressure is a maximum and at the centre of a rarefaction the pressure is a minimum.

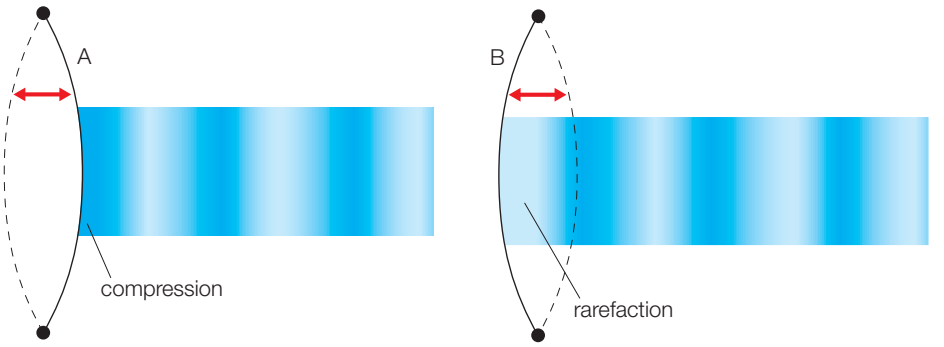


Figure 18.1 The vibrating wire produces a series of compressions (higher density of air molecules) and rarefactions (lower density) in the surrounding air.

As these compressions and rarefactions are formed by the vibrating wire, they travel outwards through the medium in all directions around the vibrating wire (figure 18.2), and not only in the plane of vibration of the wire as might be expected. Sound is said to have a 'bulk effect', which means that the effect is felt over the entire volume of medium that surrounds the vibrating wire. The sound is heard in all directions around the wire.

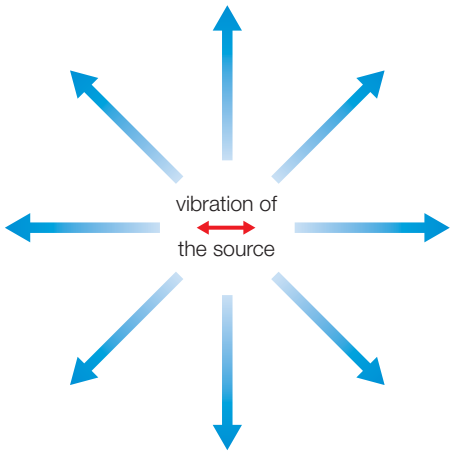


Figure 18.2 Sound travels in all directions around the source. Although the vibrations of the source are in the plane shown, the resulting sound waves travel in all planes around the source.

As we saw in chapter 16, as long as the source keeps on transmitting energy to the medium, that energy is propagated outwards by the medium. The waves travelling outwards from the source are longitudinal, since the particles of the medium are vibrating in the same direction as that in which the waves are travelling.

Compressions and rarefactions can only be formed if the movement of the medium is along the same line as the direction of travel of the wave.

Whatever the nature of the source, sound travels as longitudinal waves from the source. Since a longitudinal wave must have compressions and rarefactions, and these cannot form unless a medium is present, it follows that we cannot have a sound wave in a vacuum (see Practical activity 18.1). As you will soon see, we can measure the speed of sound by timing echoes, as in Practical activity 18.2.

Practical activity 18.1

Sound does not travel in a vacuum

You will need:

- bell-jar with an opening at the top, on a suitably greased stand
- electric bell with long connecting wires passing through a rubber bung that fits the bell-jar opening
- switch
- exhaust pump connected by rubber or plastic tubing to the base of the bell-jar
- suitable power supply
- means of hanging the bell from the bung.

Method

1 Fit the rubber bung securely into the top of the bell-jar with the bell hanging from the bung (see figure 18.3).

2 Connect the bell to the battery with the switch open.

3 Thoroughly grease the base of the bell-jar and place the bell-jar firmly on the stand.

4 Switch on the bell circuit with the exhaust pump switched off.

5 Having listened to the sound of the bell in the bell-jar, now switch on the pump (this will draw air out of the bell-jar) and continue listening to the sound.

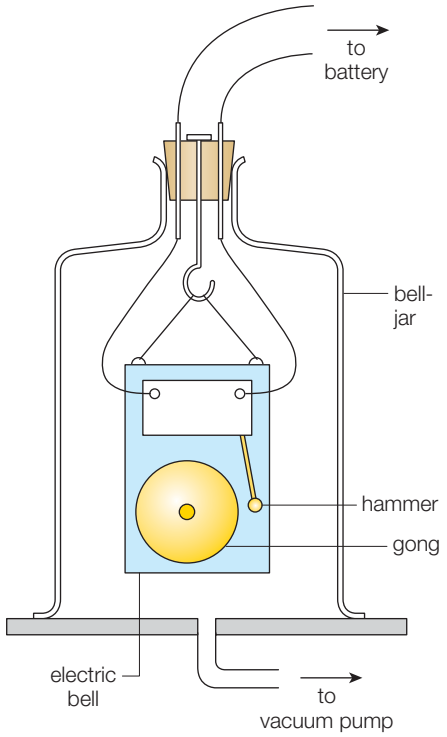


Figure 18.3

Observation

The hammer can be seen hitting the gong and the bell can be heard. However, as the air is drawn out of the jar, the bell sounds fainter and fainter. After a time the bell cannot be heard at all, even though the bell circuit switch is still on and the hammer can be seen hitting the gong.

ITQ1

Can you explain:

(i) why the pitch of the notes emitted by a steel pan depends on the area of the pan which is struck?

(ii) how a note is produced when you blow vigorously over the open end of a pen cover?

(iii) why a simple tune can be played using bottles containing water to different levels?

(iv) why a tuning fork will give out a note when struck?

Explanation

As the air is drawn out of the bell-jar, the number of air molecules in it becomes smaller. Since these molecules are carrying the energy given out by the bell, the amount of energy leaving the bell-jar decreases and so the sound becomes fainter. When there are too few molecules left in the jar to convey sufficient energy from the bell to the wall of the bell-jar, the observer no longer hears any sound.

Group discussion

If less and less energy is taken away from the bell as the bell-jar is evacuated, what effect will this have on the bell itself? Where does the energy of the vibrations go?

Inference

The presence of air in the bell-jar is crucial for the transmission of the sound.

Conclusion

Sound cannot travel through empty space.

ITQ2

Astronauts on the Moon cannot speak to each other as they might do on Earth. Explain why. How do they communicate with each other?

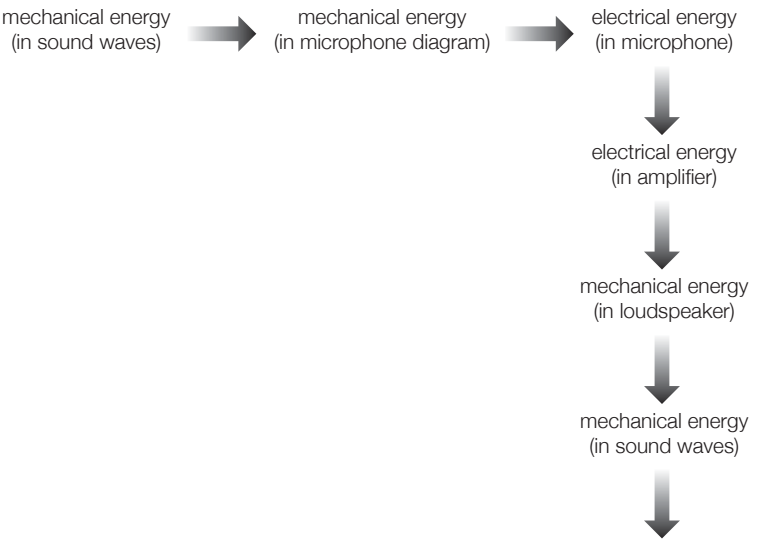
Reception of sound

For a sound to be heard, the sound waves must reach the hearer's eardrum. When this happens, the eardrum is made to vibrate with the same frequency as the waves themselves. The mechanical vibrations are passed on to the three small bones of the ear, the malleus, incus and stapes, then on to the middle ear and from there to the inner ear. There the vibrations are converted into electrical signals, which are taken to the brain by the auditory nerve. These electrical signals are then interpreted by the brain as sound.

Sound may also be received by a microphone and eventually delivered by a loudspeaker through a series of energy conversions. These conversions may be shown on a flow diagram as follows:

BIOLOGY: transmission of sound from the ear-drum

Note the process of the continuous conversion of energy from one form to another from the vibrating source to the brain.



Between the microphone and the loudspeaker there is often an amplifier, the purpose of which is to amplify the sound put out by the speaker by increasing its amplitude (as we shall see later) and therefore the energy delivered.

What about the speed at which sound travels? This can be an important issue in certain situations such as the timing of races where the duration of the race is several seconds and not minutes or hours. We take an example to show the timing of a race can be quite inaccurate because the person timing uses sound rather than the flash of the starting pistol (light) to start his watch.

When a race is timed, the timing should begin not when the sound of the starting pistol is heard, but when the flash of light from the pistol is seen. If the sound of the pistol was taken as the starting moment, the measured time could be inaccurate by the length of time for the sound to travel from the starting point of the race to the finishing point. This inaccuracy could have a very important effect on the timing of a 100 m race, as shown in Worked example 18.1.

Worked example 18.1

Estimate the percentage error and the timing error in the timing of a 100 m race in which the time was measured from the moment the sound of the starting pistol was heard by an observer at the finishing point rather than the flash of the pistol. Take the speed of sound to be 300 m s^{-1}.

Solution

The sound of the pistol will take 100 m/300 m s^{-1} to travel from the starting point to the finish, that is, about 1/3 of a second. Assuming that the race takes 10 s, the percentage error in the timing would be (0.33/10) × 100 or about 3%. In a 100 m race, 0.1 s is an appreciable time interval that can be easily detected by timing devices. It could be the degree of precision of the watch. The actual error in the time is 3% of the time for the race and that is about 0.33 s. An error of 0.33 s is three times as large as the precision of the timing device and it would be easily identified by the timing device. Such an error in timing could give rise to serious controversy.

This problem of a large timing error may be avoided by using the flash of a pistol shot to start the timing rather than the sound of the shot, the speed of light being so very much larger than the speed of sound.

Speed of sound compared with speed of light

Light travels in air at 300 000 000 m s^{-1} and sound at about 300 m s^{-1}. The latter value is in fact slightly higher than this. It depends on the temperature and pressure of the air. So light travels approximately one million times faster than sound.

Since the speed of the light of the flash is 3.0 × 10^{8} m s^{-1}, which is 10^{6} times as large as that of sound, the time for the light of the flash to reach the timer of the 100 metre race will be 10^{6} times smaller than when the sound was used as signal for the start. The error will therefore be 0.33 s × 10^{-6}, which is about 0.3 millionths of a second. It is unlikely that any timing device in use at the present time for timing races will measure times with this degree of precision. Such a tiny error can certainly not be measured by any watch or other timing device and so it will not matter and can be ignored.

We may also use the speed of sound to estimate the distance of thunder. We may assume that the thunder started when we see the lightning flash (as light travels so quickly). If we now multiply the speed of sound by the time interval between lightning flash and the sound of thunder, we will find how far away the thunder originated. As a rough guide, a 5 s interval means a distance of about 1 mile.

ITQ3

Assuming that a thunder clap is heard 5 seconds after the lightning flash is seen, estimate how far away the thunder and lightning originated.

How can the speed of sound be measured outside of the laboratory?

Although accurate methods can be used to measure the speed of sound in the laboratory, a very simple and unsophisticated method can be used outside of the school laboratory on the playing field. The method uses the convenience of the echo.

Practical activity 18.2

Measuring the speed of sound using echoes

You will need:

- two flat pieces of wood as clappers, which produce a loud noise when they are banged together
- a high, distant wall; there must be no obstructions between the clappers and the wall
- stopwatch
- measuring tape of suitable length to measure the distance from the clappers to the wall.

Method

1 Stand a long way away from the wall and strike the clappers together. Listen for the echoes reflected from the wall.

2 Adjust the rate of 'clapping' in such a way that a clap coincides with the return of the echo from the wall. Maintain this rate of clapping.

3 Measure the time taken from clap 1 to hearing the echo of clap 11 at this particular steady rate. This is a 10-clap interval.

4 Repeat this observation and find a mean time for the two measurements taken together

5 Measure the distance from the position of the clapper to the wall.

Calculation

Suppose the times for 10 clap intervals are t_1 and t_2.

The average time interval, T, between claps is $(t_1 + t_2)/20$ s. In this interval the sound from the clap has travelled from the source to the wall and back, a distance of $2d$, where d (in metres) is the distance from the source to the wall. If the speed of the sound is taken as v, then, since

$$\text{speed} = \frac{\text{distance}}{\text{time}}$$

we have $v = \frac{2d}{T}$

One serious drawback of this method is the high degree of uncertainty in the measurement of the time between successive claps. The uncertainty (and therefore the error in the measurement) is due to the size of the reaction time of the experimenter in relation to that of the time interval. They might be very close to each other.

In order to get some idea of the size of the error that could result from the reaction time of the experimenter, we take a worked example.

Worked example 18.2

An experiment like that in Practical activity 18.2 was carried out using a stopwatch capable of measuring time intervals to the nearest 1/100th of a second (such as a digital watch). The distance from the clappers to the wall was 100.0 m.

(a) What are the degrees of uncertainty in the distance from the clappers to the wall and in the time interval between claps, if the interval is taken as the average calculated from 10 intervals?

(b) Calculate the range within which the value of the speed would lie with these uncertainties.

(c) Express the result to 2 significant figures.

Solution

Assume that the tape is graduated in centimetres. You are unlikely to measure such a large distance to better than 5 cm: even finding the exact perpendicular from the wall is not easy.

Distance to the wall, $d = 100\,\text{m} \pm 5\,\text{cm}$

The maximum percentage error in the distance is therefore

$$\frac{2 \times 5\,\text{cm}}{100\,\text{m}} \times 100\% = 0.1\%$$

Although the precision of the watch is 0.01 s, the reaction error of the person using it might be as high as 0.2 s. Since it is quite possible that the watch might have been started too soon and stopped too late, we must add the reaction errors. This gives a maximum error of 0.4 s.

Because of the size of this error, we must make the time interval to be measured by the watch as large as we conveniently can. If we mistakenly use only one interval (which will be $2d/v = 200\,\text{m}/300\,\text{m s}^{-1}$, or 0.7 s, roughly), we find that the maximum error is comparable with the time for one echo to return.

If we assume that 10 intervals are used, the time for this will be 10×0.7 s, or roughly 7 s, and the maximum percentage error becomes $(0.4/7) \times 100$ or about 6%. If therefore our calculation for the speed yields a result of $320\,\text{m s}^{-1}$, the overall error will be 6% of this value.

The uncertainty in the value of v is due to uncertainties in both d and T. The latter of the two is by far the more important.

We will therefore ignore the error in the distance, d, its value being so small (0.1%, obtained by taking $(2 \times 5\,\text{cm})/100\,\text{m}$). A 6% maximum uncertainty in 320 is about 20. We would therefore expect the result to be $(320 \pm 10)\,\text{m s}^{-1}$. The limits within which the value lies would be 310 m and 330 m. The value will therefore be somewhere between 310 and $330\,\text{m s}^{-1}$. Expressed to 2 significant figures, the value would be $320\,\text{m s}^{-1}$. The significance here is rather poor. We could improve it by 'spreading the error'. In order to do this, we use 50 intervals instead of only 10 with the following result.

Using 50 intervals instead of only 10 would give the same reaction error of 0.4 s but in a total time that will be 5 times as large or 35 s. This would produce an uncertainty of 1/5 of the previous error (of 6%) which is a little more than 1%, in which case the maximum uncertainty in the result for the speed would be 1% of $320\,\text{m s}^{-1}$, that is about 3 or $4\,\text{m s}^{-1}$. The result would now lie somewhere between 318 and $322\,\text{m s}^{-1}$. Thus by spreading the reaction time error for starting and stopping the stopwatch over a large number of intervals we have been able to reduce significantly the effect of the reaction error and obtain better precision and therefore significance.

The two ranges of uncertainty are shown in figure 18.4.

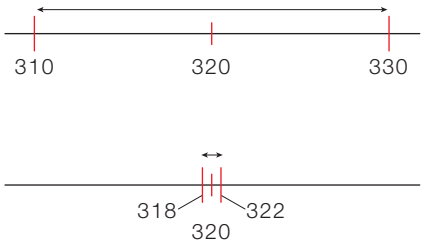


Figure 18.4 Range of uncertainty in determining the speed of sound using the echo method. The uncertainty depends on the time measured.

ITQ4

Work out the percentage error and the limits of uncertainty if 100 intervals were used in Worked example 18.2.

The effects shown by sound are those shown by all waves whether they are transverse or longitudinal, and whether they are matter waves or not. The first effect we will discuss is reflection

Reflection of sound

We begin with reflection. You will learn in chapter 21 that reflection is that effect where the energy of a wave falling on an interface between two media of different densities 'bounces back into the medium from which it is coming. Like almost all phenomena in physics, the effect is governed by laws called 'the laws of reflection'. Most of the energy in a sound wave which falls on a solid wall will therefore be reflected, because the wall is very much denser than the air in which the incident sound rays are travelling at first and the wall is also rigid. If these conditions of a higher density and rigidity are not met, then some of the energy falling on the wall called the 'incident energy' could be absorbed at the interface.

Like other waves, sound waves obey the laws of reflection (page 41). An investigation to see whether the laws hold can be carried out with the arrangement shown in figure 18.5.

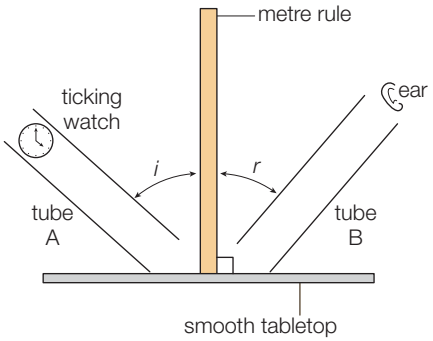


Figure 18.5 The reflection of sound.

CHAPTER 21

Two cardboard or (better) plastic tubes with narrow cross-sections are arranged as shown with the lower ends resting on a smooth table. A small ticking watch is placed just inside the upper end of tube A. The experimenter keeps one ear close to the upper end of the other tube, B, and holds tube B at different angles until the ticking of the watch sounds loudest. At that point, a large protractor is used to measure the angle between the tubes and a metre rule held perpendicular to the table where the tubes meet. If the angles which the tubes make with the rule are denoted by i and r as shown in figure 18.5, then at the point where the ticking is loudest, it is found that

$i = r$ (within the limits of error of i and r).

To investigate whether the equality of the angles i and r is within the limits of experimental error, you will have to consider the precision of the protractor used to measure them, remembering to add the errors at both ends of the angle being measured.

Angle i is called the 'angle of incidence' and angle r the 'angle of reflection'. Stated in words, the law states that

Echoes are the most familiar example of the reflection of sound. It should be clear from a consideration of the first law that, for echoes of a sound to be heard at its source, the surface of the reflector must be perpendicular to the incident sound ray. If it is not, the sound will be reflected, but not towards the source.

When a sound ray is reflected from a surface,
angle of incidence = angle of reflection

This is, of course, the first law of wave reflection. It does not apply only to sound waves, but to all waves, whatever the type, transverse or longitudinal, matter or electromagnetic.

The second law of reflection states that the incident ray, the reflected ray and the normal at the point of incidence are all in the same plane. This can be checked by placing a large drawing board on the table to touch both tubes along all their lengths. If the metre rule is found to be parallel to the board, then the sound rays and rule are in the same plane.

Refraction of sound

We consider next the interesting phenomenon called refraction. In order to understand this effect we must (i) know what refraction is and (ii) know the conditions required for it to occur. It can occur not only for sound, but for all types of wave energy.

What is 'refraction'?

Refraction is that effect in which a ray of wave energy in leaving one medium to enter another where the speed of the wave is different, changes course (that is, bends) on entering the second medium.

As an example we take sound entering water from air. A swimmer who is under the surface of water in a swimming pool is listening to music coming from a device in an overhanging branch of a tree. The speed of sound in air is 300 m s^{-1} and in water (believe it or not) is much higher at 1500 m s^{-1}. Draw a probable path of the sound ray when it enters the pool. Clearly, there are two possible paths the ray can take (see figure 18.6). It can take either path (a) or path (b). The correct path is given by a law called **Snell's Law**.

Will the sound ray follow path (a) or (b)?

Snell's Law ›

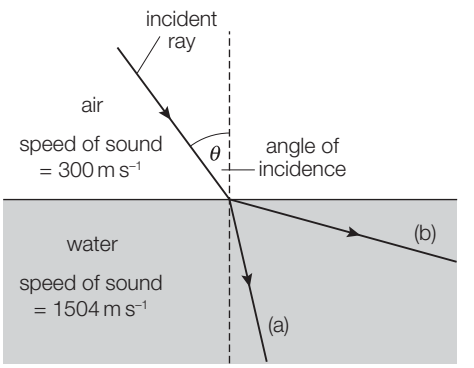
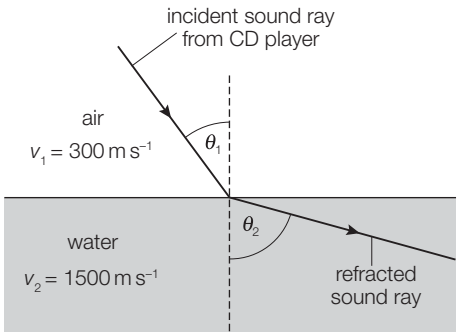


Figure 18.6



By Snell's Law,

$$\frac{v_1}{\sin\theta_1} = \frac{v_2}{\sin\theta_2}$$

$$\therefore \sin\theta_2 = \frac{v_2}{v_1} \times \sin\theta_1$$

$$= \frac{1500}{300} \times \sin\theta_1$$

so $\sin\theta_2 = 5 \times \sin\theta_1$

$\therefore$ The angle of refraction (in water) is larger than the angle of incidence (in air)

Figure 18.7

To state Snell's Law we use figure 18.7 in which the sound ray is entering the water at an angle of incidence, θ_1, and leaving the interface at an angle of refraction, θ_2. If the velocities of the sound in the two media 1 and 2 are v_1 and v_2 respectively, then the law states that

$$\frac{v_1}{\sin\theta_1} = \frac{v_2}{\sin\theta_2}$$ (Snell's law – Law no. 1)

This law can be stated as follows:

The sine of the angle between the ray and the normal in a particular medium is proportional to the speed of the ray in that medium.

We can now return to the question posed above and obtain an answer, (a) or (b), to the question. Since the ray left air (where the speed is much less) and is entering water (where the speed is very much higher), then from the law, the sine of the angle the ray makes with the normal in air will be less than the sine of the angle it makes in water.

It must also be pointed out, too, that

The incident ray and the refracted ray will be on opposite sides of the normal. (Law no. 2)

Solution to the question: will the sound ray follow path (a) or (b)?

Referring back to figure 18.6, we conclude that the correct path of the ray on entering the water will be (b), where the angle which the ray makes with the normal is larger, since the sine is larger. This solution tells us that when sound leaves air for water it bends away from the normal. If we use the values of the velocities of the sound in air and in water, we obtain

$$\sin\theta_2 = 5\sin\theta_1$$ (see figure 18.7)

MATHEMATICS: trigonometrical functions

Remembering that for angles between 0° and 90° sin θ increases as θ increases for angles less than 90° we therefore infer that the larger sine will be associated with the larger angle and the angle of refraction θ_2 is larger than the angle of incidence, θ_1 as figure 18.7 shows.

You must compare this behaviour of sound energy with that of light energy when we come to it in a later chapter. We can now discuss the refraction of sound in a more general way.

Everyday effects of refraction of sound

It is not easy to show in a school laboratory that sound waves can be refracted. However, there is ample natural evidence that sound waves are refracted. Have you ever noticed that sounds produced at ground level can be more easily heard during the night than during the day? This is explained by refraction.

The speed of sound waves in air is affected by the temperature of the air; the waves travel faster at higher temperatures. (Can you find out why? The internet might be of some help.) When sound waves encounter layers of air at different temperatures, their speed alters and they are refracted in accordance with Snell's Law.

At night temperatures of the air near the ground are lower, and the air becomes warmer the higher up we go. Thus the sound waves travel faster as they go higher, and they bend more and more away from the normal as they get higher until eventually they might well turn back and approach the earth again. Further, since their speed increases with height, and speed is proportional to wavelength, then their wavelength also increases with height. Remember, too, that the distance separating adjacent wavefronts is the wavelength. This

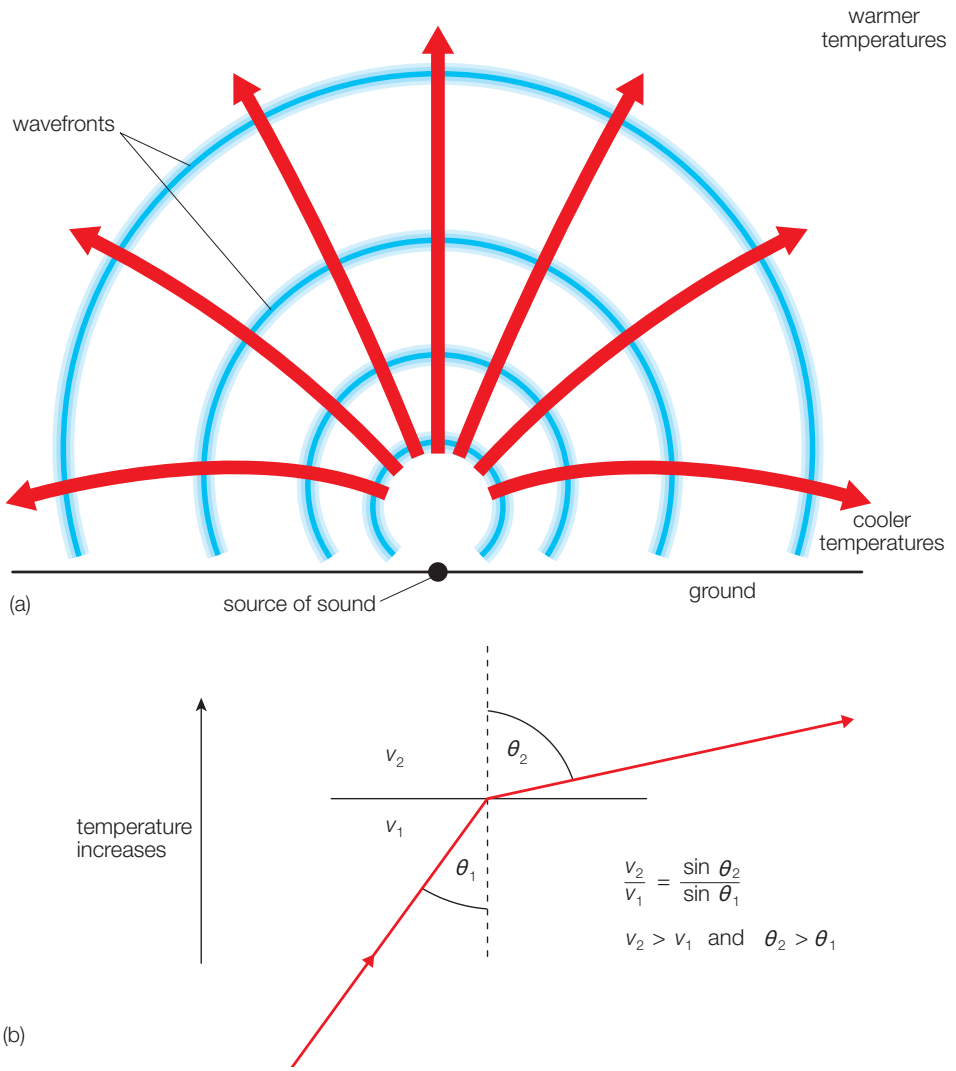


Figure 18.8 At night (a) the sound wavefronts travel faster when they move from the cooler air nearer the ground through the warmer layers of air higher up and (b) the sound 'rays' are bent away from the normal as they travel upwards.

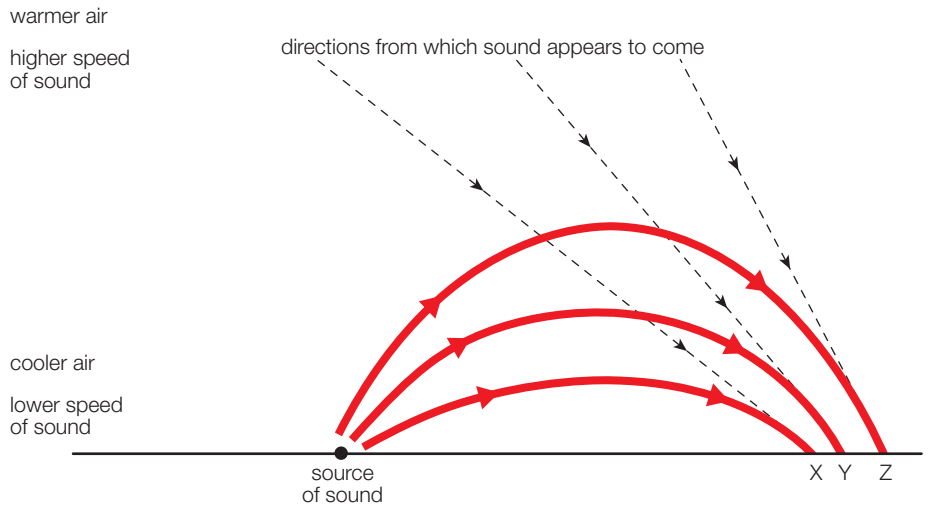


Figure 18.9 The distant sound is heard more easily at night. The sounds seem to come from above. Can you say why from the diagram?

leads to the conclusion that in the vertical plane shown in figure 18.8 the wavefronts will look as drawn. The direction of travel of the sound energy is shown by the sound rays (bearing arrows) in the figure corresponding to these wavefronts; the sound rays near to the ground after a time bend in such a way that they return towards the ground and the energy they carry can be detected. As a result of the sound waves returning towards the Earth, sounds from a distant source may appear to persons at X, Y and Z of figure 18.9 to come from the directions of the dashed straight lines ending at these points.

During the day the layers of air near the ground are hotter and those higher up are cooler. Since the temperature falls with height, so does the speed of the sound waves; they slow down. Therefore the sound waves bend towards the normal as they rise. Figure 18.10 shows the directions of the sound waves from a source at ground level. Have you noticed that they curve upwards? Can you draw the wavefronts for these rays? How do they look compared with those at night? Do you see why the sounds at the source would not be heard some distance away?

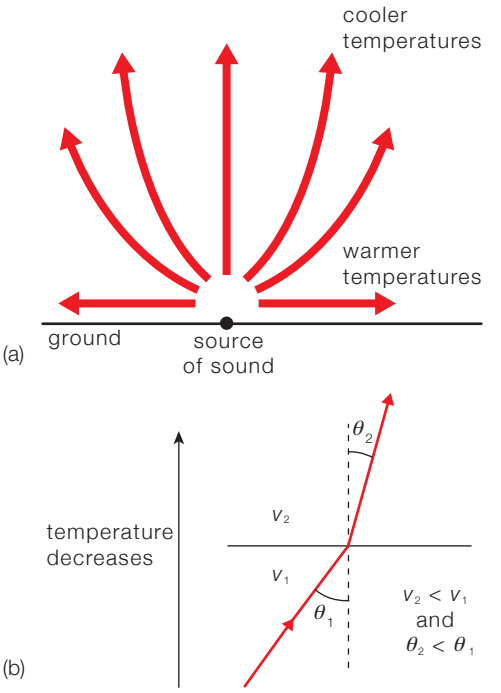


Figure 18.10 (a) During the day, air temperatures are higher nearer the ground. As the sound waves move higher, they are refracted away from the ground. (b) They are refracted towards the normal.

Interference of sound waves

In order to discuss interference of sound waves, we must, as was the case in refraction, first understand what it is and what are the conditions required for it to take place.

We sometimes see beautiful colours on the ground in wet streets where petrol or motor oil has spilled. This is caused by an effect called interference, which is one of the effects shown by waves. As sound is a wave motion it is therefore also capable of showing interference. To understand interference, one must have a good grasp of the meaning of phase difference and of the principle of superposition.

Superposition: two vibrations interacting with each other

Two sources at S_1 and S_2 in figure 18.11 are vibrating in the same direction and generating transverse waves. Suppose that:

(i) two of the waves travel along the paths S_1X and S_2Y, meeting at P, and

(ii) the phase relationship between them does not change, then this implies that

(iii) the frequencies of the waves leaving the two sources must be the same at all times.

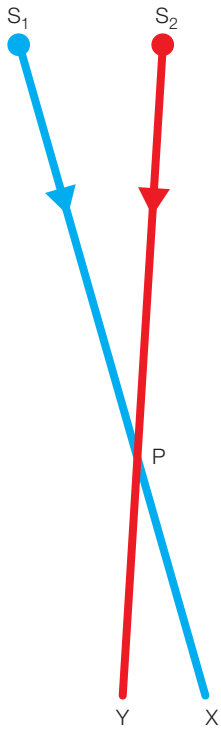


Figure 18.11 Two travelling waves crossing at P and interfering.

coherent sources ›

If the second and third conditions are met, the sources S_1 and S_2 are said to be **coherent sources**. We must know at the outset that this condition of being coherent must be met if the two waves are to interfere, that is, produce a resultant vibration at the meeting point whose features will not change. We will see that the amplitude of the resulting oscillation there depends on the phase difference between the two vibrations taking place at P where the two waves meet.

When the vibrations at P are in phase

Suppose that the waves have the same amplitude, and also that when the waves meet at P they both have the same phase. Since the displacements of the two vibrations are along the same transverse line, we add the displacements to get the overall displacement. There will therefore be an increased maximum displacement or amplitude at P. So:

> **When two vibrations meet in phase, the resulting vibration will have an amplitude equal to the sum of the individual amplitudes.**

ITQ5

A dipper in a ripple tank is vibrating up and down. The distance between its highest and lowest points is 3 cm. What is the amplitude of the vibration of the dipper?

superposition ›

We say that when the waves meet at P they are superposed. The principle of **superposition** states that, in general:

> **Whenever two vibrations taking place along the same line are superposed, the resultant displacement will be the sum of the individual displacements at every stage of the vibrations.**

For transverse waves, displacements (and amplitudes) above the mid- or equilibrium position are usually defined as positive and those below have a negative sign. With longitudinal waves, displacements (and amplitudes) to the right of the equilibrium position are often defined as positive and those to the left as negative.

Figure 18.12 (a)–(g) shows the changes of displacement that take place at P during half a cycle of the waves' progress, beginning with a maximum displacement for each wave. As the waves pass through P, the resulting displacement at P at any stage is twice the value for each wave. The amplitude is therefore twice that for each of the separate waves.

Figure 18.13 (a) is the displacement–time graph for the resulting vibration of a particle at P when the vibrations are in phase.

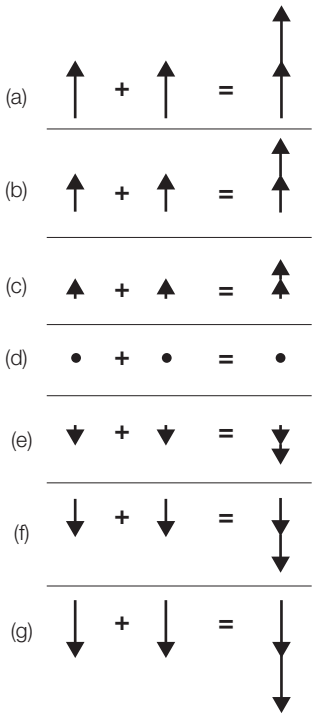
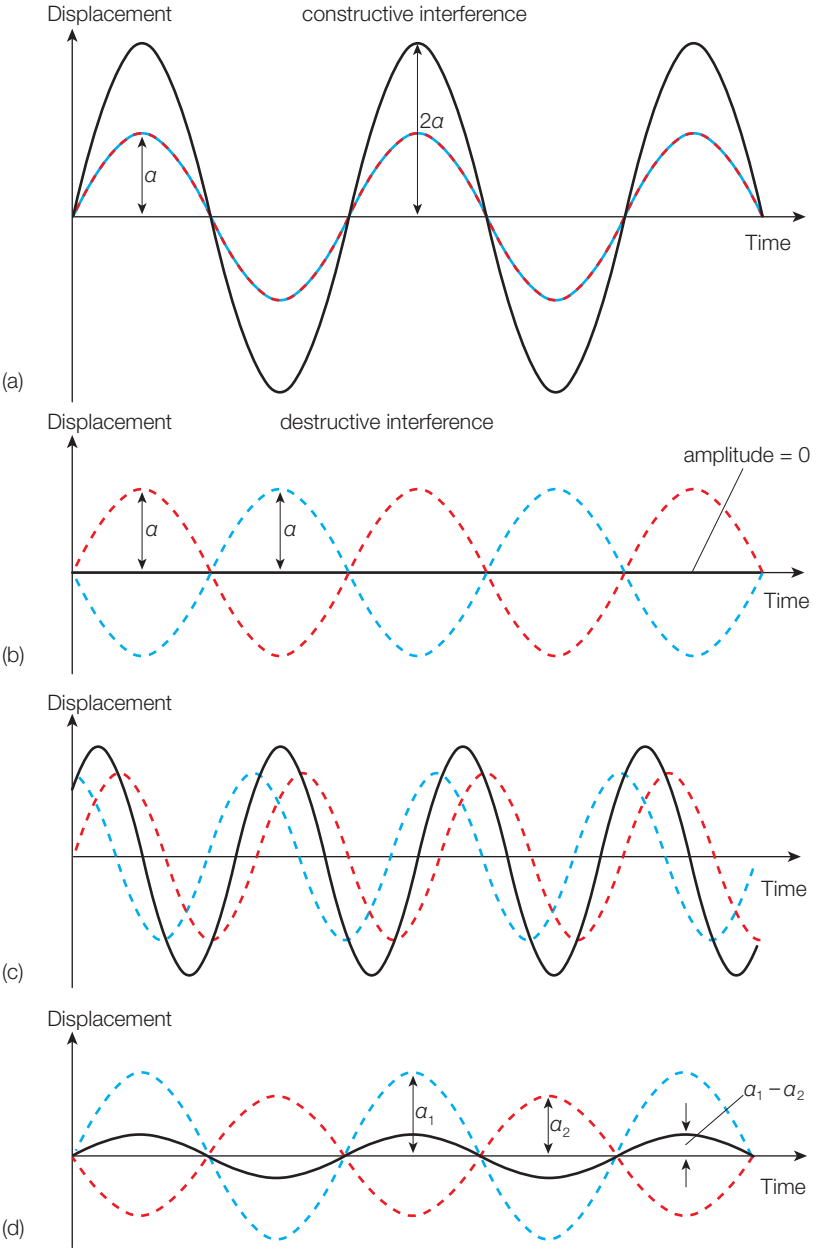


Figure 18.12 (a)–(g) The net displacement at P when two waves meet. The displacement at any moment is the sum of the separate displacements. As the vibrations have the same amplitude and are in phase, the amplitude at P is twice the amplitude of the individual waves.



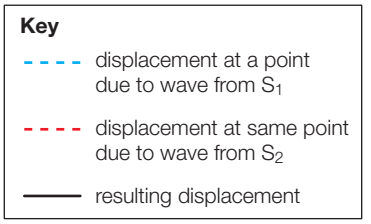


Figure 18.13 Displacement–time graphs for particles at different points P where waves meet. (a) Waves of the same amplitude, and with the same phase at P. The amplitude of the resulting vibration is the sum of the amplitudes of each of the individual vibrations, i.e. $2a$. These vibrations produce constructive interference. (b) Waves of the same amplitude, at a point where the vibrations are directly out of phase: their displacements cancel each other out. Hence there is no displacement of particles at that point and the resultant amplitude is zero. These vibrations produce destructive interference. (c) Example of waves of the same amplitude, at a point where the vibrations are out of phase but not directly out of phase: the resulting amplitude is between 0 and $2a$. (d) Waves of unequal amplitude, with vibrations directly out of phase: the resulting amplitude is the difference between the amplitudes of the individual waves, i.e. $a_1 - a_2$.

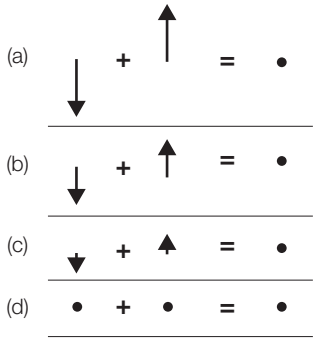


Figure 18.14 (a)–(d) When two vibrations of equal amplitude are directly out of phase, the resultant amplitude is always zero.

When the vibrations at P are directly out of phase

Suppose that the waves produced by sources S_1 and S_2 again have the same amplitude but are directly out of phase (opposite phases or 'in antiphase') when they arrive at P. The displacements will then always cancel each other out and the resulting amplitude will always be zero (figure 18.14 (a)–(d)).

Figure 18.13 (b) shows the displacement–time graph for a particle at P in the case where the vibrations are directly out of phase. Particles at that point are never displaced from their equilibrium position.

Suppose that the waves are of equal amplitude and the vibrations are out of phase but not directly out of phase. The resulting amplitude at P will have a value somewhere between 0 and $2a$. Figure 18.13 (c) shows an example of this.

Figure 18.13 (d) shows what would happen if the vibrations were directly out of phase but one of the vibrations has a greater amplitude than the other. The resulting amplitude would be the difference between the two amplitudes. We must remember that all that has been said about superposition in the context of transverse waves would apply equally to longitudinal waves. If we regard the displacements as vectors, then to obtain the resultant displacement at any time or in any position, we simply add together the individual displacements.

We are now in a position to discuss interference of sound waves.

Demonstrating the interference of sound waves

Sound waves, like any other type of wave, will show all of the effects that arise from the principle of superposition such as interference and diffraction. While it is not easy to find interference of sound waves taking place 'in nature' or as an everyday occurrence as one might interference of light waves – think of coloured oil films on water, coloured soap bubbles and the like – it is not so for sound interference. You might think that two large speakers on the stage of your school auditorium might be able to produce interference of the sound produced by the person addressing the audience on 'Speech Day' but, sadly, this is not likely, as the conditions required for the effect to be produced are so stringent! It is hardly likely, therefore that we will come across interference as an everyday occurrence. However, we can create the right conditions for the effect to be produced, as explained below.

To demonstrate interference, we need two small identical sources of sound, such as two tweeters (very tiny loudspeakers – the sources must be tiny), connected up so as to produce sound waves in phase (and therefore coherent). We also need a reliable means of detecting sound, either a sensitive ear (upon which one cannot rely, since hearing is very subjective and what person A might say he/she can hear person B might not hear at all) or, better, a microphone connected to a cathode ray oscilloscope (see chapter 30 for the mode of operation), which can 'pick up' the sound, faint though it may be, and convert it into electrical oscillations.

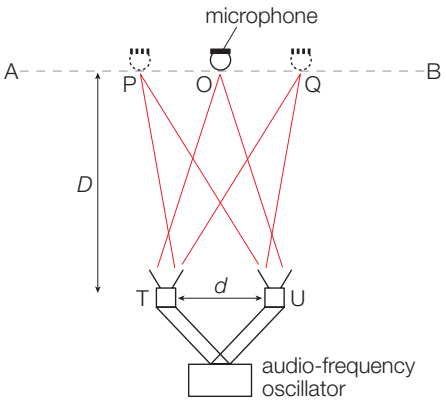


Figure 18.15 Demonstrating interference of sound waves.

The apparatus is arranged as shown in figure 18.15. The two tweeters, T and U, separated by a distance d (about 0.5 m), are each connected to the same audio frequency oscillator (AFO). This means that the sound waves emitted by the tweeters always have the same frequency and phase relationship – that is to say, they are coherent. Also, the volumes of the sound, that is, the amplitudes of the waves emitted by the tweeters, will be very similar.

The frequency of the audio frequency oscillator is set at about 1 kHz. The microphone is moved slowly along an imaginary line parallel to the speakers and about 3 m (D) from them (line AB in figure 18.15). A vertical trace of varying height is seen in the centre of the oscilloscope screen. The height of the trace varies between a maximum value (figure 18.16 (a)) and zero (figure 18.16 (b)), with (a) corresponding to a case of constructive interference where

An audio frequency oscillator (AFO) is a piece of electronic equipment which can produce electronic oscillations (in voltages) of frequency ranging from about 10 Hz up to about 20 kHz, which is the range of frequencies to which the human ear is sensitive. When a loudspeaker is connected across the output terminals of an AFO and switched on, a sound of frequency corresponding to the setting of the instrument is heard.

the waves are meeting in phase, and (b) to the case of destructive interference where the waves are meeting exactly out of phase.

As the microphone is moved along AB, the path difference between the wave-trains it detects from the two tweeters will vary continuously. Thus at O in figure 18.15, the two wave-trains have travelled the same distance (from symmetry) and have no path difference and consequently no phase difference. They interfere constructively and give (at O) a vibration of the air molecules with a large amplitude (amplitudes added). This is constructive interference. The energy of this vibration is converted by the microphone to electrical energy (voltage and current). The voltage, oscillating in its turn, causes the electrons hitting the oscilloscope screen to move up and down, giving a vertical trace (figure 18.16 (a)). The same thing happens at P and at Q, where the path difference is one wavelength of the sound wave. This is found again for other positions beyond P and Q (which are both the same distance from O) where the path difference is a whole number (greater than one) of wavelengths and so constructive interference occurs at these places.

At one position between P and O and another between O and Q, the line on the oscilloscope screen becomes a dot or possibly disappears altogether. At these places there is little or no vibration of air molecules. The path difference here is one-half of a wavelength, and the phases of the interfering wave-trains will be directly opposite, giving an overall resultant at all times of zero, assuming, of course, that the interfering displacements are of equal size. There is therefore destructive interference at these positions.

To listen directly for the interference effects, the experimenter's ear should be at the same horizontal level as the tweeters, so that the wave-trains will be travelling in a horizontal plane. In this case one hears a rise and fall in the volume of the sound as one moves parallel to the line of the speakers. However, the person whose ears can 'pick up' the rise and fall of loudness of the sound as he or she moves from O to P or from O to Q will have to possess extremely sensitive and very sharp hearing.

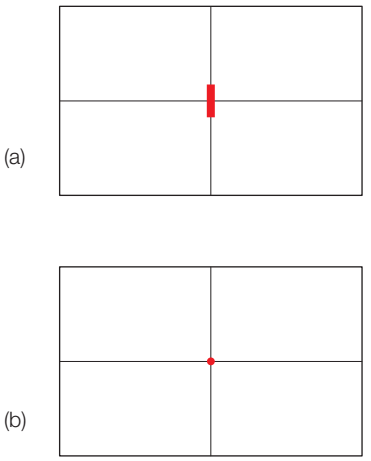


Figure 18.16 Oscilloscope traces: (a) where the sound is loudest (constructive interference); (b) where the sound is least loud (destructive interference).

Diffraction of sound

diffraction ›

Diffraction is the sideways spreading of wave energy as waves go past an edge or through a narrow opening. Figure 18.17 shows a wave passing through such an opening. The energy carried by the wave is found to extend into areas at the sides of the 'straight-through' wave. The theory of this effect is beyond the scope of our syllabus, but the effects are what we should be acquainted with.

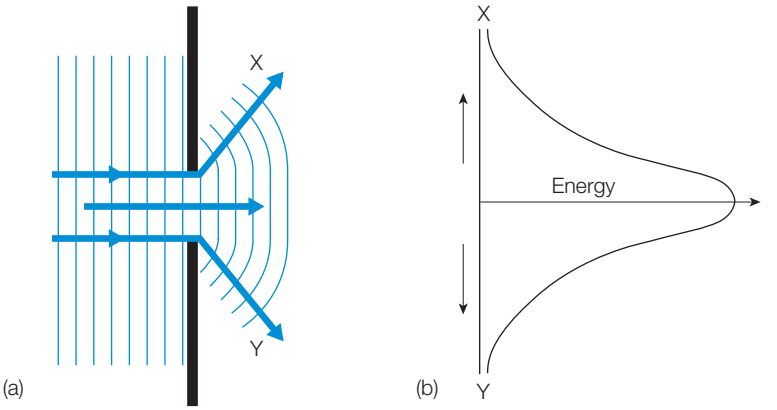


Figure 18.17 Diffraction of plane waves at a narrow opening. (a) The waves spread sideways as they pass through the narrow opening. (b) How the energy is distributed in the diffracted beam. The further from the centre of the diffracted beam, the less the energy diffracted.

Diffraction at an opening or narrow gap

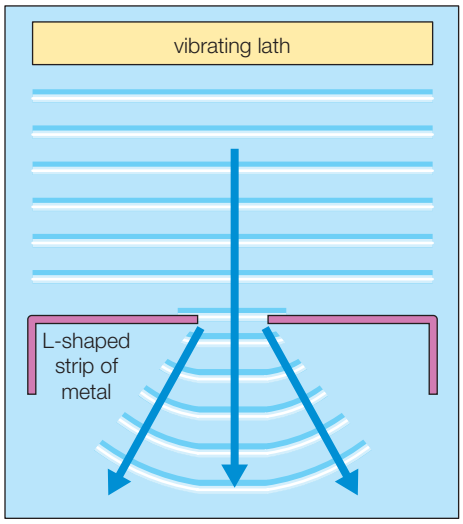


Figure 18.18 Using a ripple tank to investigate diffraction. The wavelength of the waves, λ, and the width of the opening, *d*, can be varied. As the ratio λ/*d* increases, the amount of diffraction (or spreading) increases.

This effect may be demonstrated using water waves in a ripple tank (figure 18.18). The plane waves are produced by a straight lath dipping in and out of the water and the narrow opening is formed by two L-shaped strips of metal standing opposite each other.

Plane wavefronts advancing parallel to the opening spread sideways when they go beyond it. It is important to know that the factors that determine the amount of spreading (or the angle of diffraction) are:

1 the width of the opening, *d*, and

2 the wavelength of the wave, λ.

In order to change the wavelength of the incident waves when using the ripple tank, we alter the current in the motor. The smaller the current, the lower the frequency of the waves produced and the greater their wavelength.

The amount of diffraction is determined by the ratio

$$\frac{\text{wavelength of the waves}}{\text{width of the opening}} = \frac{\lambda}{d}$$

Increasing the wavelength while the width of the opening remains constant

Figure 18.19 shows the effect of increasing the wavelength of the wave as the width of the opening remains constant. In figure 18.19 (a), the wavelength, λ, is much less than *d*, the width of the opening. No diffraction can be detected, since the width of the beam past the opening does not change. The angle through which the rays at the edge of the beam are deviated from their course (the angle of diffraction) is very small indeed.

When the wavelength is made larger (figure 18.19 (b)) but with the width of the opening kept the same, so that the wavelength is comparable to the width of the opening, there is now some noticeable spreading, that is, more diffraction of the incident beam. The ratio λ/*d* is larger, and the angle of diffraction, λ, increases.

In figure 18.19 (c) the wavelength is equal to the width of the opening, i.e. the ratio λ/*d* has a value of about 1, and the angle of diffraction is nearly 90°.

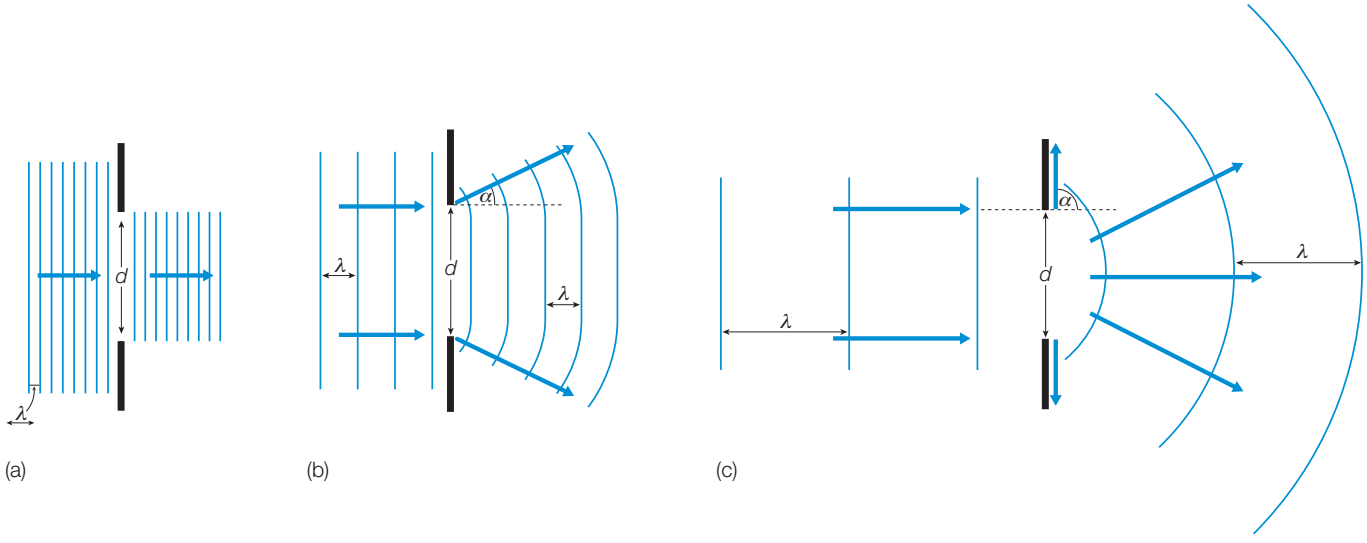


Figure 18.19 As the ratio λ/*d* increases, so does the amount of diffraction. (a) The opening is large compared with the wavelength (λ/*d* is small) and no diffraction can be detected. (b) λ/*d* is larger, and there is noticeable diffraction at the edges of the beam. The angle of diffraction is a. (c) The value of the wavelength is equal to the width of the opening. There is maximum diffraction, the wavefronts are circular and angle α has increased to about 90°.

Diffraction at an edge

Diffraction also occurs when a wave goes past a single sharp edge. The amount of diffraction in this case depends mainly on the wavelength of the waves, the larger the wavelength, the greater the degree of diffraction. Thus, whereas sound waves (longitudinal) with wavelengths of the order of tens of centimetres are considerably diffracted at the edge of a building at a street corner, light with wavelengths about one million times smaller will not be diffracted at the edge of a razor blade. Figure 18.20 illustrates diffraction at an edge for two waves with widely different wavelengths. In figure 18.20 (a) some diffraction of the waves takes place because the wavelength is large: the wavefronts bend round the edge. In figure 18.20 (b) hardly any diffraction occurs because the wavelength is smaller.

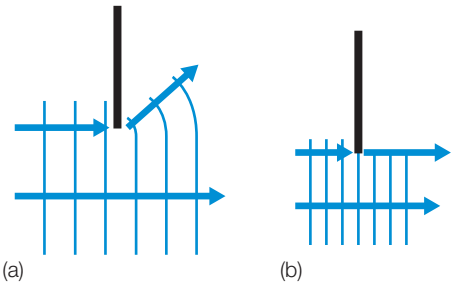


Figure 18.20 Diffraction at an edge. (a) Waves of large wavelength are diffracted. (b) There is little diffraction with smaller wavelengths.

We can now see to what extent diffraction can be considered an everyday occurrence and one that can happen incidentally. We begin with a simple demonstration.

Close all the windows and doors of your laboratory. Ask someone to speak normally in the laboratory from a position directly behind the closed door, but a few metres away from it as shown in figure 18.21. Ask someone else to listen at various positions, A, B and C, outside the room and to report on whether he or she can hear what the person inside is saying.

Usually you will find that not much is heard. Now repeat the experiment with the door open. You should find that the speaker in the room is heard clearly from position A (and you will know why), and again from position B but not quite so clearly as from A (can you now explain why using what we have just discussed about diffraction?) and, possibly too, from position C but still less clearly. (Why do we say 'possibly'?)

Nothing is heard when the door is closed, suggesting that no sound is passing through the walls or through the windows or door when these are closed.

Could the following be some of the answers to the questions just posed?

- When the door is open, sound is passing through the doorway without obstruction. The sound is heard at A because the waves pass straight through the open doorway.
- The sound is heard at B because sound waves reach the listener after diffraction at the near edge of the doorway.

The sound may also be heard at C because of diffraction. Under what condition(s) may the sound be heard more certainly at C? The answer will involve mention of the width of the doorway and the wavelength of the sound waves created by the person in the laboratory.

The results described above are only obtained if the wavelength of the sound of the speaker's voice is about the same size as (but smaller than) the width of the door. If the speaker was now replaced by someone with a much higher pitch (for example, a young child), it is possible that nothing would be heard at C, since the wavelength of the sound emitted by the child would now be much smaller than the door width. (Check this!) Would this make the ratio λ/d smaller or larger (see page 277)? What would be the effect of moving to a longer room, or an area such as the opening at the end of a long corridor? What difference would all this make?

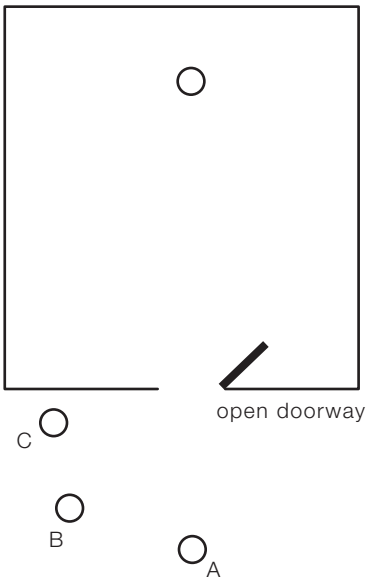


Figure 18.21 Sound can be diffracted round an open doorway. Where will the sound be faintest? Why?

Group discussion

1 Why is it important for the person who creates the sound to be at the back of the lab and not near the open door?

2 Why should it matter if the person in the lab is a child or an adult?

3 Why is the position of the person listening outside the laboratory so important to hearing the sound?

4 Assuming that the conditions are 'right' for persons A, B and C all to hear the sound coming through the open door, which one of them will hear:
- the loudest sound, and why?
- The faintest sound and why?

ITQ6

When it is otherwise quiet in the early hours of the morning, you might distinctly hear the bass (lower) notes of the music at a party some distance away. However, you don't hear any treble (higher) notes at all. Explain this.

There are a number of other everyday observations that strongly suggest that sound is diffracted. One example with which you may be familiar is that sounds from a band in one street can be heard in other streets, even though the sources are out of sight. For another example consider ITQ6.

Pitch and loudness of sound

pitch ›

Two features of sound that may help to distinguish one sound from another are **pitch** and loudness. We generally can tell a child's voice from that of an adult and the voice of a male from that of a female. What we use is the pitch of the voice and what we mean by this is the frequency level of the voice.

We say that the pitch of a soprano singer's voice is higher than that of a bass singer. What is it that distinguishes the bass voice from the soprano? The answer lies in the range of frequencies of the notes in the two voices. Whereas the bass singer may produce notes of frequencies ranging from 80 Hz up to about 350 Hz, the soprano's range may be from 250 Hz to 1 kHz. So it is the frequency of a note that determines its pitch. We may therefore classify voices according to increasingly higher range of frequencies as:

low pitch				*high pitch*
bass	baritone	tenor contralto	mezzo-soprano	soprano
80–350 Hz	110–425 Hz	200–700 Hz		250–1000 Hz
(male voices)			(female voices)	

A note will be louder if more energy goes into producing it. If a guitar string is plucked or a steel pan is struck vigorously we would expect the note emitted to be loud (conservation of energy?). We would also find that the amplitude of vibration of the string or the vibration of the pan was large. (How would you detect the latter?). Louder sounds are associated with waves and vibrations of larger amplitude. There must be a connection between the energy associated with a sound and the amplitude of the oscillations that produce the sound. The amount of sound energy falling on the eardrum of a person per second will determine how loud the note sounds or the loudness of that sound. Loudness is an aural impression produced when the eardrum oscillates with large amplitude. The loudness is not the same for all those hearing the sound, however. It is subjective and so we should not trust our ears to judge how much energy is present in a sound.

The audible range

It is well known that the human ear will only respond to sounds that are within a certain range of frequencies. This range is generally taken to be:

	frequency range/Hz
• For music:	50–15 000
• For speech:	100–8000

Clearly, then, since ultrasound frequencies lie outside of these ranges, humans cannot hear ultrasound.

Ultrasound

The term 'sound' may be applied to longitudinal mechanical vibrations that travel in solids, liquids and gases. The speeds of these waves depend on the nature and temperature of the medium in which they travel. Sound vibrations of frequencies up to about 15 kHz will produce the sensation of hearing when

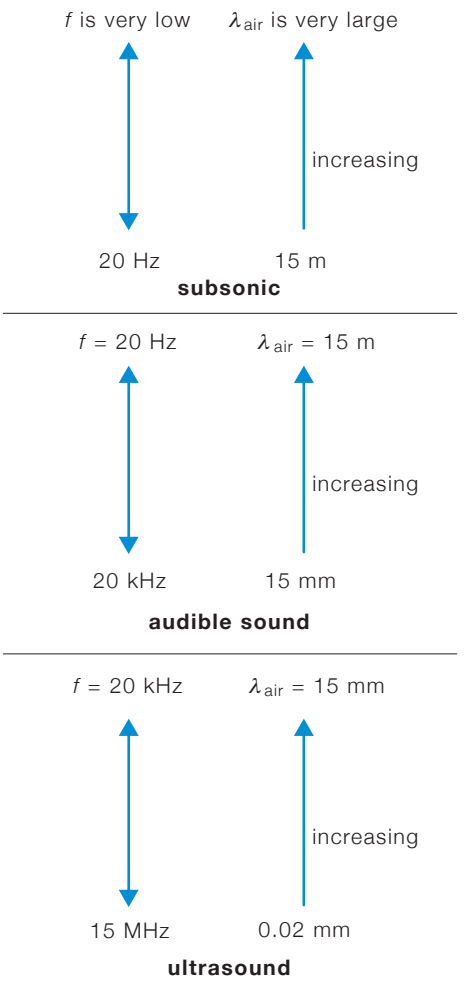


Figure 18.22 The ultrasound spectrum.

ITQ7

Calculate the wavelength of ultrasound in water if its speed is 1500 m s^{-1} and its frequency is 500 kHz.

ITQ8

A depth-sounder on a boat sends out a signal and receives an echo 0.5 s later. If the speed of sound in water is 1500 m s^{-1}, how deep is the water?

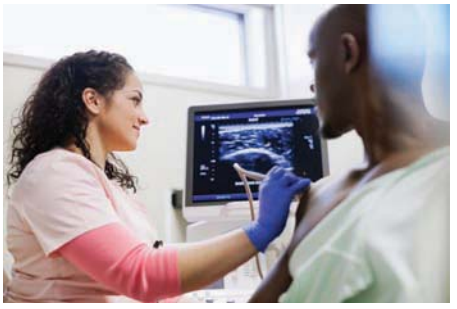


Figure 18.23 This patient is having an ultrasound examination.

they fall on the human eardrum. Higher values than this are called ultrasound, since the frequencies concerned are beyond those of normal hearing. (The Latin word *ultra* means 'beyond'.) The frequency and wavelength spectrum for ultrasound is shown in figure 18.22.

Ultrasound behaves in the same way as ordinary sound. The effect that is most often used, however, is that of reflection. It would be extremely difficult to produce the other effects practically, since the wavelengths involved are so very small.

Ultrasound is used to determine the depth of bodies of water. The frequencies used in depth-sounding are high, at about 50 kHz, and there are good reasons for this. The high frequency means that the wavelength is very short, so that there is little diffraction or spreading of the ultrasound wave energy. (Remember that the ratio λ/d determines how much diffraction will occur – the smaller λ can be made, the less the spreading of the energy.)

This means that a beam of ultrasound energy can be made more directional; it can be 'aimed' much better than ordinary sound.

In addition, because the wavelength is small, the ultrasound waves are reflected from smaller objects or irregularities on the seabed rather than being diffracted by them, and so more detail can be seen. Remember that in reflection the information (borne by the waves) returns to the sender, whereas in diffraction the energy does not return to the sender. We say that using a smaller wavelength improves the 'resolution'. It better enables the experimenter to detect tiny objects.

A further advantage of using ultrasound for this purpose is that the ultrasound is not confused with other sounds in the water since the detector does not respond to ordinary sound frequencies.

Another very important application of ultrasound is in medical imaging. It is also used in medicine very routinely nowadays to provide images for diagnoses and for monitoring (figure 18.23). You will be aware that the gender of the foetus of an unborn child can now be predicted with the use of ultrasound imaging.

The advantage of ultrasound imaging over X-ray imaging is that, unlike X-rays, it does not destroy or damage tissue, as far as we know.

To produce the ultrasound images, ultrasound pulses are sent into the patient's body and are reflected at the interfaces between sets of tissue of different 'ultrasound densities'. Remember that, for reflection of the waves to take place, there must be a difference between the densities of the media on the two sides of the reflecting interface. For example there is a marked difference between the ultrasound densities of amniotic fluid surrounding the foetus of the womb and the foetus tissue and so the reflection at the interface between these two is strong and good images can be obtained. The time taken for the transmitted pulse to return to the detector (the echo) is a measure of the distance of the interface from the source. The picture built up from the echoes is used to estimate the position, size and shape of, for example, internal organs, tumours or cysts. It seems that one of the most important medical uses of ultrasound nowadays is providing images of developing foetuses.

Ultrasound is used to remove grease and dirt from metallic objects immersed in a cleaning fluid. The ultrasound vibrations are passed through the fluid and help to shake off the dirt particles without harming the metal surface in any way. It is also used in industry to detect the presence of cracks in metals.

You might be aware that bats very successfully use ultrasound signals of frequencies around 150 kHz to find their way around without bumping into objects. (A bat can fly unharmed through the blades of a spinning electric fan, it is believed.) Dolphins and whales also use ultrasound of similarly high frequencies in order to navigate.

Chapter summary

- All sound waves consist of longitudinal vibrations.
- All sound waves need a medium for propagation.
- The speed of sound waves varies with the medium, but in any one medium its value depends on the temperature.
- When sound undergoes reflection, it follows the laws applying to waves, that is:
  angle of incidence = angle of reflection
- Sound waves may undergo refraction when they leave one medium for another of a different density and in which they have a different speed.
- When sound waves undergo refraction at an interface between two media, the ratio of the sines of the angles of incidence and refraction equals the ratio of the respective speeds of the wave in the two media, which gives
  $$\frac{v_1}{\sin\theta_1} = \frac{v_2}{\sin\theta_2}$$
- Sound waves can be made to interfere under suitable conditions.
- Sound may be divided into at least two categories namely:
  (i) ordinary sound of frequency about 50 Hz to about 15 kHz (can be heard by humans)
  (ii) ultrasound of frequency about 20 kHz to about 10 MHz (can be 'heard' by bats, whales, etc.).
- A sound or note can be characterised by pitch and by loudness.
- The pitch of a note is related to the frequency of the note.
- The loudness of a note depends on the amplitude of the waves transmitting the note.
- Ultrasound has a number of uses in industry and in medicine.

Answers to ITQs

ITQ1

- The frequency of the note emitted depends on the area and thickness of the part of the steel pan which is struck. These differ over the surface of the pan.
- Blowing over the open end of the pen cover sets air in the cover into vibration. The frequency will depend on the length and diameter of the cover.
- The volume of air in the bottle determines the frequency of the sound emitted when the air is set into vibration. Different frequencies can therefore be produced from the bottles from which a simple tune can be composed.
- When struck, the prongs of the fork vibrate and this vibration produces notes the frequencies of which depend on the length and thickness of the prongs of the fork.

ITQ2 The astronauts cannot talk normally with each other because there is no atmosphere on the Moon to carry the energy. No waves are produced, therefore, when they speak, they communicate by radio.

ITQ3 1500 m

ITQ4 The percentage error is now about 0.6%. The maximum error is 2 m s^{-1}. The limits of uncertainty are ± 1 m s^{-1} and the result would be between 319 m s^{-1} and 321 m s^{-1}.

ITQ5 1.5 cm

ITQ6 The bass notes have lower frequencies and therefore longer wavelengths than the treble notes. In order to be hard, the sound will have to pass around buildings and also between them, that is, be diffracted around them. Sound of longer wavelengths is more likely to be diffracted than that of the lower wavelengths, since the value of the wavelengths will be closer to the sizes of the spaces involved (10 m to 20 m).

ITQ7 3 mm

ITQ8 375 m

Examination-style questions

1 As a longitudinal wave motion, sound requires a medium for propagation. Would you expect air at high pressure to transmit sound faster than air at lower pressure? Give a reason for your answer.

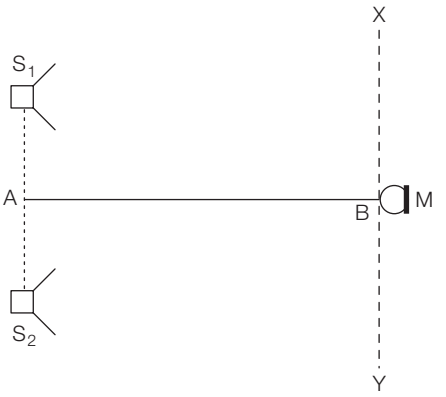
2 A boy wishes to demonstrate interference of sound waves to his class. He decides to use two large loudspeakers in his school playing field. He therefore sets up the two loudspeakers separated by about 50 m and connected to a CD player through an amplifier. He switches on his CD player, which plays the West Indies cricket anthem.

(i) Outline the method he should use to find out whether or not he is obtaining interference of the sound from the two speakers.

(ii) How would he know whether his experiment is successful?

(iii) Do you think his experiment is likely to be successful? Give a reason for your answer.

3 S_1 and S_2 are two small, identical speakers supported on retort stands at the same height above a long table. AB is the perpendicular bisector of the line joining the two speakers, as shown in the diagram. A microphone, M, is mounted at B, the midpoint of XY, and is connected to a cathode ray oscilloscope (CRO). The time-base circuit of the CRO is switched off. The speakers are now connected in parallel with the output terminals of an audio frequency oscillator, which emits a note of constant frequency.



(i) Describe and explain what will be seen on the screen of the CRO when the audio frequency oscillator is switched on.

One of the speakers is now disconnected from the audio frequency oscillator.

(ii) Describe and explain what is now seen on the screen of the CRO.

The disconnected speaker is now reconnected and the microphone is moved along the line XY towards Y.

(iii) Describe and explain what is observed on the screen of the CRO as the microphone is moved.

One of the speakers is again disconnected and the microphone moved along XY from B as before.

(iv) Describe and explain what will now be seen on the CRO screen as the microphone is moved.

5 (i) Calculate the wavelength of ultrasound travelling in soft tissue if the frequency of the sound is 5 MHz and its speed in the soft tissue is 1550 m s^{-1}.

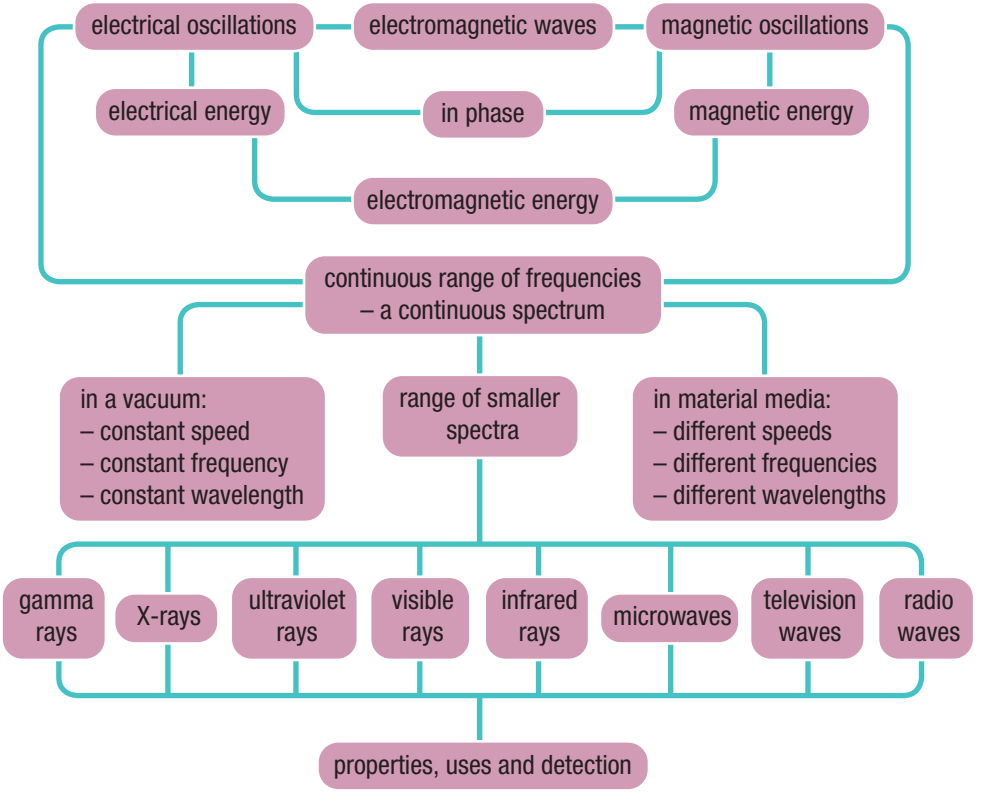
(ii) How many times as long will the wavelength be if the frequency were reduced to 50 kHz?

(iii) Hence describe the advantages of using high-frequency ultrasound for ultrasound scanning in medical diagnosis instead of low-frequency ultrasound.

19 Electromagnetic Waves

By the end of this chapter you should be able to:

- describe the circumstances in which electromagnetic waves are produced
- describe the nature of electromagnetic waves
- recall that all electromagnetic waves travel at the same speed in a vacuum
- recall that the speed of electromagnetic waves in a medium varies with the nature of the medium
- recall that the temperature of the Sun is several thousands of degrees Celsius
- recall that the electromagnetic waves emitted by the Sun have a very wide continuous range of frequencies and therefore a continuous range of wavelengths
- recall that this continuous range of frequencies (called the continuous electromagnetic spectrum) can be subdivided into smaller continuous spectra, each with a characteristic range of frequencies and wavelengths
- remember that these different types of spectra do not have sharp boundaries but overlap with their neighbours
- recall that the waves of each of these smaller spectra, if they do interact with matter, show particular physical effects when they do
- recall that the radiation in each of these smaller spectra is produced in a particular way
- understand that because of these effects the different types of electromagnetic radiation have particular uses



Introduction

Electromagnetic waves are that type of wave motion among which occurs the particular type which enables us (perhaps most importantly) to see. They also are the type that brings radio and television programmes to our homes, and make plants grow. We also use them to examine tissues in the human body and the range of applications to which they have been put is steadily getting wider. That this is so is probably due to the fact that the range of electromagnetic wavelengths is enormous: the wavelengths of radio and television waves are very large (they can be hundreds of metres), whereas those that are used to examine tissue are extremely short, being millionths of a micron. How, though, do these waves differ from those we generate on a string or on a slinky, and how do they carry energy if they can travel in empty space?



How large is one micron in metres?

In this chapter we shall discuss the nature, properties and uses of the various forms of electromagnetic waves. We shall see that they do not depend on a medium for transmission and that, like water waves, sound waves and all other types of waves, they do show wave effects such as interference and diffraction and, importantly, like all other types of waves, they, characteristically, transmit energy from one place to another.

How are electromagnetic waves produced?

You will remember how waves were produced on a rope and generated in a ripple tank. An object was moved to and fro to produce vibrations in the medium around the source. The vibrating object was continuously changing speed (or accelerating) and direction. Electromagnetic oscillations are generated in much the same way. They are formed when moving electric charges are made to change speed or accelerate. Thus the oscillating electrons in the transmission masts of transmitting stations are the source of electrical and magnetic energy that flows away from the mast in the form of waves, in much the same way as mechanical energy flows away from the source, carried by water waves or waves on a slinky. This applies to all forms of electromagnetic waves, whether they are as short as the very shortest, gamma waves or as long as the very longest, radio waves.

In the material or matter waves discussed in earlier chapters, the energy was transmitted by the medium in the form of kinetic and potential energy in particles of the medium. There was a continuous transformation of energy between potential and kinetic forms. These two forms of energy, which are associated with mass, are called mechanical energy. Electromagnetic waves do not need matter to take their energy forward; they can travel just as well in a vacuum as in any suitable material medium such as glass, air or plastic. How, then, is their energy, often called **electromagnetic radiation**, transmitted or propagated?

electromagnetic radiation ›

The nature of electromagnetic waves

Like all other waves, electromagnetic waves consist of vibrations of something. For sound waves it is the vibration of molecules of the medium, for water waves the vibration of small bodies of water. What is it that vibrates in the case of electromagnetic waves? It cannot be anything material, since these waves can travel in a vacuum.

We made mention in Chapter 4 of electrical forces and electrical energy associated with them and also of magnetic forces and magnetic energy associated with them and you will learn in later chapters of electric fields and magnetic fields associated with these forces and these energies. For the

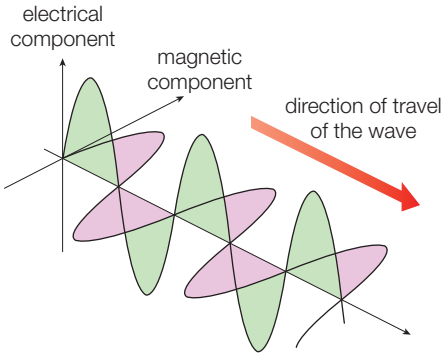


Figure 19.1 An electromagnetic wave has an electrical component and a magnetic component travelling together along the same path, but not in the same direction. Both components are transverse, and the directions in which they vibrate are perpendicular to each other. The two vibrations take place in phase and the direction of travel of the wave is at right angles to the plane containing these two directions.

purposes of this chapter we will regard electric fields and magnetic fields vibrating just as we did material particles of the medium, water or slinky, that is, the 'thing' which vibrates to propagate the wave. In the same way that, in these particles, potential energy is constantly being exchanged with kinetic energy, in a very similar way electrical energy in the electromagnetic wave is constantly being exchanged with magnetic energy as the wave moves along its path. Unlike material waves, however, where both energies are found in the vibration of one 'thing', a particle of the medium, in one plane, in the electromagnetic wave the energies reside in two different vibrations, one an electrical vibration and the other a magnetic vibration. These vibrations take place in perpendicular planes and are always in phase. The direction of travel of the electromagnetic wave is perpendicular to the plane containing the directions along which the two vibrations take place. You can see a representation of these two waves, the electrical one and the magnetic one, in the diagram of figure 19.1.

The waves that leave the Sun are produced as a result of the behaviour of electrically charged particles in it (which are the vibrators) that causes them to give out energy. This energy is of two different, but closely connected, kinds: electrical and magnetic. You will learn later (in section D) that moving electrical charges produce magnetic effects. The behaviour of the charges in the Sun leads to electromagnetic energy being emitted.

electromagnetic energy ›

It is this **electromagnetic energy** that is carried by the waves, the electrical energy component by one set of waves, and the magnetic component by the other – hence the adjective 'electromagnetic' to describe the combined energy. Be reminded that both of these sets of waves are transverse, but the vibrations of which they consist take place in planes that are perpendicular to each other. Be reminded, too, that the two components, one magnetic in nature and the other electrical in nature, and both travelling along the same path, but in perpendicular planes, are in phase at all times. They together form an electromagnetic wave (figure 19.1).

Naming the different 'types' of waves in electromagnetic radiation

The Sun is at a temperature of several thousands of degrees Celsius and because of this, many different kinds of electromagnetic waves leave it. The wavelengths of these waves form a series, from extremely small at one end and progressing steadily and continuously (that is, without a break) to values of the order of hundreds of metres at the other. Because of the continuous nature of this series of wavelengths, the range of wavelengths is described as a **continuous spectrum** of electromagnetic radiation. We may now consider the series (the continuous spectrum) to be divided up into 'bands' of wavelengths, these bands being based on common properties and behaviour and effects of the radiation in particular circumstances. Each of these bands is a spectrum in its own right. It represents a collection of continuous electromagnetic radiation which all behave similarly, each with its own name that describes it. Thus we have the following bands starting from the very shortest and proceeding to the longest: gamma rays; X-rays; ultraviolet (UV) rays; visible rays; infrared (IR) rays; micro waves; television and radio waves.

continuous spectrum ›

Waves that all have the same effect on matter are put into one category and given one name and the set of bands or spectra together make up the Sun's complete electromagnetic spectrum. For example, the group of waves that have a warming effect on matter that absorbs their energy are called 'heat waves' or, more formally, 'infrared rays', the noun 'rays' being used only for historical reasons. These waves together form the infrared spectrum. X-rays are a different 'kind' of radiation from the infrared, only because they have a quite

different effect on matter into which they might be directed. They can give rise to images of the matter they enter. They therefore have a different name, X-rays, 'X' because of the circumstances surrounding their discovery. Their set of radiation is called the 'X-ray spectrum'. Radio waves and microwaves such as those used in telecommunications have different effects in very different circumstances. So, in short, each band of waves that forms a 'family' is given a special name and, as we shall see, has a particular function and produces a particular effect.

It is much easier to identify a band of radiation having the same effect on matter or showing similar behaviour by their wavelengths in air rather than by their frequencies. We would prefer to say that the wavelengths of light waves are from 400 nm to about 700 nm, rather than the frequencies of light waves are 'from … to …'. So types of radiation are generally identified by their wavelengths. The unit of wavelength used is most often the 'nanometre' which is 1.0×10^{-9} m. Sometimes (for very short waves) the 'micron' is used, one micron being 1.0×10^{-6} m.

Electromagnetic spectra

Electromagnetic waves from the Sun



Make a guess at the temperature of the surface of the Sun.

As stated above, because of its extremely high temperature, the Sun is able to emit waves covering a very wide range of frequencies. All of these waves travel at the speed of light in a vacuum, that is $3.0 \times 10^{8}\,\text{m s}^{-1}$. It is interesting to note that the waves at the low wavelength end of the spectrum are very, very short, indeed, but at the long wavelength end are not at all so very long. Thus, whereas the shortest waves, gamma waves, may be as short as 1.0×10^{-14} m, the longest waves, radio waves are only about 10^{6} m long. It seems then that the Sun emits a very wide range of high-frequency or low-wavelength electromagnetic waves.

Worked example 19.1

The wavelength of a wave emitted by the Sun is 600 nanometres. Calculate the frequency of the wave.

Solution

The symbol c is used to represent the speed of electromagnetic waves in free space or a vacuum.

Using the equation $c = f\lambda$, we have

$$f = \frac{c}{\lambda}$$

$$= \frac{3.0 \times 10^{8}\,\text{m s}^{-1}}{600 \times 10^{-9}\,\text{m}}$$

$$= 5.0 \times 10^{14}\,\text{Hz}$$

The frequency of 5.0×10^{14} Hz is very high indeed; it corresponds to visible light with a colour between yellow and green. However, the Sun also produces waves of very much higher frequencies, such as gamma rays and X-rays. They could be of the order of 10^{23}!

The complete electromagnetic spectrum

The complete spectrum of electromagnetic waves extends from the shortest gamma rays to the longest radio waves. All of these waves belong to the same big family called the complete electromagnetic spectrum. This complete, extensive and continuous spectrum is divided up into smaller (that is,

visible spectrum ›

narrower) spectra each one showing properties which are peculiar to it. For example, the **visible spectrum** consists of that band of radiation whose wavelength extends continuously from around 400 nm right up to about 700 nm. One has to say 'about', since there is no sharp dividing line between neighbouring spectra. So no one knows where visible radiation ends on the low wavelength side and where ultraviolet begins on the high wavelength side, or where X-radiation ends on the low wavelength side and gamma radiation begins on the high wavelength side. We might say therefore that the different spectra of electromagnetic (e.m. for short) radiation 'merge' into neighbouring spectra. The boundary between neighbours is blurred.

Components of the full or complete e.m. spectrum are shown in figure 19.2. You will notice that neighbouring spectra do show an overlap. One cannot say exactly where one spectrum ends and the next begins. The name given to a wave in an overlapping region seems to depend on how the wave was produced. For example, a wave in the region between gamma rays and X-rays might be called an X-ray wave (if it was produced by an X-ray machine), or a gamma-ray wave if it was emitted by a radioactive nucleus (see chapter 35).

ITQ3

A radio station is known as '100 FM', indicating that its programmes are carried by radio waves of frequency 100 MHz. What is the wavelength of the radio waves which take its programmes?

CHAPTER 35

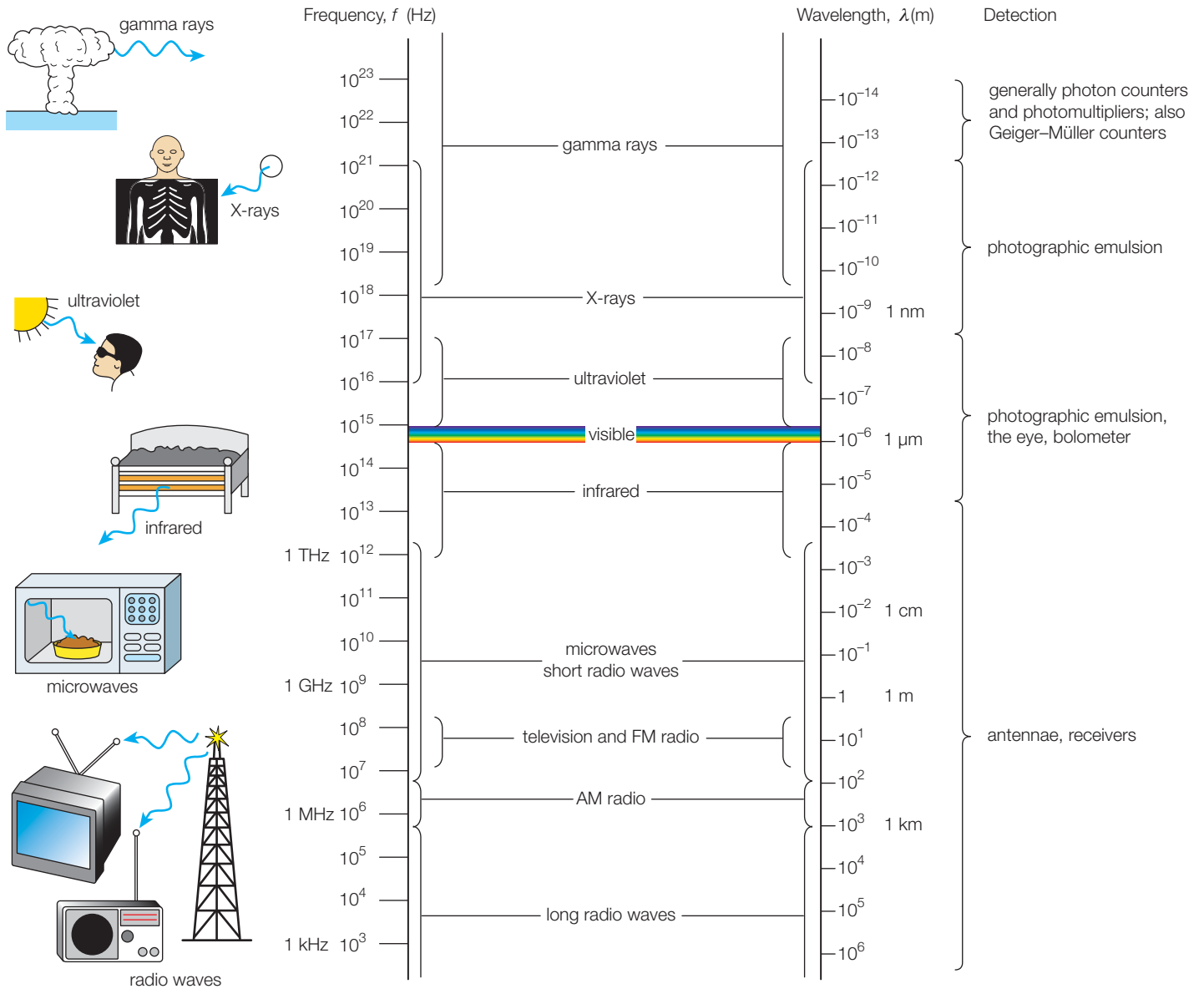


Figure 19.2 The electromagnetic spectrum.

Although the Sun produces electromagnetic waves of a very wide range of frequencies, it does not emit waves of low frequency. (Could it be that its temperature is too high for this to be possible?) Low frequency e.m. radiation is mostly created by artificial means, usually by electrical circuits. The first such circuit was used in 1889 by Heinrich Hertz, the German physicist, after whom the unit of frequency was named. Electrical circuits can be used to produce electromagnetic waves of wavelength 1×10^{-3} m (microwaves) to 1×10^{4} m (long-wavelength radio waves). The wavelengths for television broadcast waves are between those of microwaves and radio waves.

The use of the word 'rays' rather than 'waves' for some parts of the electromagnetic spectrum, for instance X-rays, is merely historical.

Can you find out the history of the use of 'rays' instead of 'waves'?

Infra is Latin for 'below'. Thus infrared waves are those whose frequencies are just below those of red light in the spectrum. Similarly, *ultra* is Latin for 'beyond', and ultraviolet light has frequencies just higher than (beyond) those of violet light in the spectrum.

Special properties and uses of electromagnetic waves

We will begin with the very shortest radiation and the type whose frequencies are staggeringly high, gamma radiation.

Gamma rays (or γ-rays)

The phrase 'gamma rays' is used to mean 'a stream of gamma ray photons'

These rays carry large amounts of energy because of their high frequency. Because of this, they are highly penetrating; several inches of lead may be needed to absorb gamma rays.

They are dangerous to living tissue and they affect photographic plates. (Look up 'the gamma camera'!). Gamma rays are used for sterilising foodstuffs and medical equipment, for medical diagnosis and for detecting cracks in metals.

You will soon be meeting the word 'photon' in connection with electromagnetic radiation. The photon is the smallest quantity of e.m. energy that can be obtained from radiation of a given frequency. The size of this varies directly as the value of the frequency. Thus, the radiation with the highest frequencies (e.g. those of gamma rays) will have the highest energies.

Gamma rays are emitted by radioactive nuclei. You will read about this in a later chapter.

Notice how wide the gamma ray spectrum is – very much wider than most of the other spectra!

X-rays

This is the radiation of the next shortest wavelength after gamma rays. They are similar to gamma rays; their photons do have a lot of energy for the same reason that gamma rays are so energetic. They can therefore penetrate flesh and bone tissue (figure 19.3) but not lead. Unlike gamma rays which are only produced naturally (from radioactive emission, a natural process) X-rays are produced artificially by using an X-ray machine. It is well known that they are widely used for medical diagnosis and therapy, for treating certain skin disorders and for studying crystal structure. This is another very wide spectrum. Note the wide overlap with the gamma ray spectrum, its neighbour.



Figure 19.3 An X-ray image of a human joint.

Group discussion

Do you know which scientist first 'hit upon' X-rays accidentally and why they were then (and still are) called 'X' rays?

Ultraviolet rays

The prefix 'ultra' was used in the last chapter when describing sound whose frequency was above those associated with ordinary sound. The prefix means 'above'. For the same reason this form of e.m. radiation is named ultraviolet because the range of frequencies associated with its radiation are above the frequencies we associate with violet. Again there is quite an overlap with X-rays, its 'higher' neighbour. There is no overlap at all, however, with the visible spectrum whose range is the very smallest of all the spectra. Any overlap here would be impossible. The distinction between ultraviolet (UV, for short) radiation and visible radiation is therefore quite sharp.

Some animals (mostly birds and insects) can 'see' ultraviolet light, although humans cannot. These waves are absorbed by glass. Although a photon of this radiation does not have as much energy as a photon of X-rays or gamma rays, the radiation can nevertheless damage tissue, can ionise atoms and, as is widely known, cause sunburn. They promote chemical reactions, and are used in medicine, in the detection of forgeries and to reveal 'invisible ink'. It is obtained from very hot bodies and, of course, from the Sun.

Visible rays

These should need no introduction. Without the availability of this form of radiation we cannot see objects. What is very significant about this spectrum is the fact that the range of wavelengths that characterise it is extremely narrow compared with even the least wide of the others. The range of wavelengths is from about 400 nm to about 700 nm – a mere 300 nm.

Visible radiation is refracted by glass and other transparent media. The eye and optical instruments use this property to form sharp images. Light rays travel in straight lines but can be carried by optical fibres along a path of any shape. Visible radiation is essential for photosynthesis and plant growth. It is used in chemistry to identify certain elements, and in light beams and lasers for communication. It is obtained from hot sources, the hotter the source, the wider the range of visible radiation present in the radiation emitted.

An important point not to be missed is that each of the 'colours' that we take to comprise the visible spectrum itself has its 'spectrum' in that there is a small 'spread' of wavelengths that we can associate with it. It is a case then of having a mini yellow spectrum or a mini red spectrum, as it were, within a larger visible spectrum. This means that no one colour has only one wavelength, but a continuous, narrow spread of wavelengths, a narrow spectrum. As expected, we really cannot tell exactly where one colour changes to the next, where, for instance yellow ends and orange begins.

Infrared radiation

Infrared is so described because the frequency band it occupies is below that of red (*infra* is Latin for 'below') in frequency terms. The wavelength band it occupies is therefore above that occupied by red. (You do remember that wavelength varies inversely as frequency, since the speed of e.m. radiation is constant). Thus, since the red wavelengths are round about 700 nm, infrared starts from here and goes right up to about 10^6 nm or 1 mm – the spectrum is quite wide, much wider than that of light, but not as wide as that of gamma rays or X-rays.

Infrared radiation produces heating when its photons are absorbed and this is why we call the radiation heat radiation. This is its chief characteristic. Warm objects, including our bodies, also emit infrared radiation. In fact all bodies above 0 K will give off infrared radiation. For this reason, infrared radiation is used for taking photographs in hazy and foggy conditions, the infrared

rays behaving just like light rays in forming images which can be fixed on photographic plates. It is much used in satellite photography since the presence of clouds does not seriously interfere. This type of radiation is obtained from very hot sources, the hotter the source, the greater the quantity and range of infrared radiation given off.

Terahertz waves (THz waves)

This is a form of infrared radiation that is just at the high (wavelength) end of the infrared spectrum.

Terahertz radiation of frequencies just below the infrared range is emitted by most objects. Although relatively new, it has been identified as having particular and very convenient properties and it seems to offer good prospects for security technology for the future. Figure 19.4 reproduces an article that gives an insight into these prospects and possibilities. The wavelength range of THz radiation if from 0.1 mm to 1 mm.

ITQ4

'Tera' is a prefix indicating '× 10^{12}'. Calculate the frequency of Terahertz radiation of wavelength 0.5 mm. Is the use of the terminology 'Terahertz' for this radiation justified?

Instant imaging device gives GPs safe new window into the body

By Robert Matthews

A camera that can see through clothes, skin and even walls without X-rays has been developed in what is being called one of the first great technological breakthroughs of the 21st century.

The "terahertz" camera, still in prototype form, is under rapid development by scientists in Oxfordshire. It is likely to have many applications, ranging from medical scanning to identifying concealed weapons on airline passengers.

Unlike X-rays, it does not expose patients to potentially harmful radiation. Instead, it detects a form of ultra-high-frequency, or terahertz, energy waves naturally emitted by all objects.

Nor does it require people to walk through a special scanner: anything that comes within range of the terahertz camera is exposed to its penetrating gaze. Dr Chris Mann … said: "These [terahertz] waves are just below infrared energy and are given off by virtually everything around us. They are also able to pass through windows, paper, clothing and, in certain instances, walls."

While the existence of terahertz waves has long been recognised, the technology needed to capture them by camera has so far been prohibitively expensive and complex.

Earlier this year, however, the European Space Agency decided to try to build the world's first terahertz camera to allow satellites to monitor the Earth through thick cloud.

…

The first historic image taken by the device revealed the outline of a hand, clearly visible despite being hidden under a thick book.

…

The prospect of low-cost and completely safe medical imaging could provide the opportunity for every GP's surgery to have such a device. Professor Laurie Hall, an authority on medical imaging at Cambridge University, said: "It's a completely new window into the human body."

Figure 19.4 Potential uses of terahertz radiation: extract reset from the *Daily Telegraph*, 6 October 2002.

Microwaves

Microwave radiation is next in line for size of waves and has become very familiar in recent times through their use in microwave ovens and in communications technology. The wavelength range is from about 1 mm to about 1 m.

Microwaves currently find application in many diverse areas of technology, perhaps the best known being in cooking or warming food and in telecommunication. Other applications not so well known are in the areas of radar, radio astronomy and spectroscopy. Whereas the longer wavelengths are used in radar, those around the middle of the spectrum find application in warming and cooking and those that are smaller still in communication

and navigation. When used for heating food, the method works best if the food has a high proportion of water. They are a very good source of heat for this purpose, since they penetrate into the food and heat it uniformly throughout. The microwaves used for cooking and warming food are between 5 cm and 10 cm long. Microwaves are produced artificially by devices called 'magnetrons'.

Radio and television waves

These waves are probably as well known as any other in the electromagnetic spectrum. As seen below, they can be diffracted around hills (can you say why?), and so they are much used for radio and television transmission. The waves at the lower end of the spectrum are also used for radar detection of ships, aircraft and missiles.

Our radios and television sets use them for transmitting information. Their wavelengths in the spectrum are the longest, ranging from about 1 m to about 1 km or more. Within this spectrum the shortest waves serve the needs of FM radio and television and the longer wavelengths the needs of AM, MW and LW radio.

Look up the meanings of FM, AM, MW and LW, since you may not be familiar with the abbreviations and may find them baffling, since people hardly listen to radio nowadays. You should, however, find this bit of internet research interesting.

ITQ5

Why would homes situated in hilly areas be better served with long-wave radio than with FM or AM radio?

General properties of electromagnetic radiation

We saw in chapter 18 that waves show reflection, refraction, diffraction and interference and that they travel in straight lines unless diverted. Electromagnetic waves are no exception, and some examples are given below

X-rays will show diffraction if they are allowed to fall on an array of atoms whose separation is of the same order of size as the wavelength of the rays. Diffraction of X-rays by atoms has been used to determine the structure of crystals (figure 19.5).

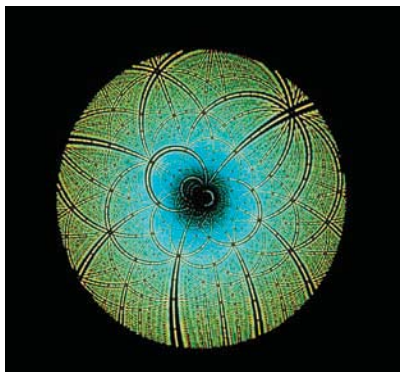


Figure 19.5 A 'Lauë' photograph as used in X-ray crystallography.

Visible rays (light) will show diffraction and interference (again in appropriate conditions). Experiments to show this for water waves were discussed in chapter 18.

Microwaves too can be made to show interference, and also diffraction, as shown in figure 19.6. As the microwave receiver (detector) is moved from side to side, the milliammeter reading (which is a measure of the intensity of the diffracted beam) varies. The reading is highest when the receiver is in line with the transmitter (source) and the opening.

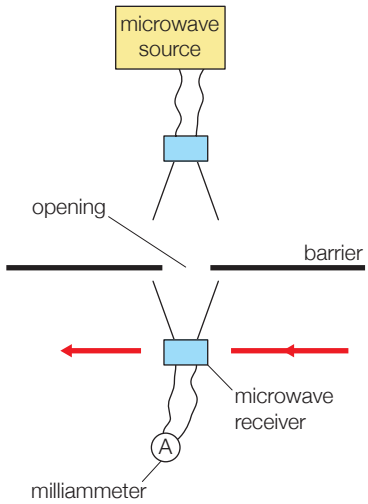


Figure 19.6 Demonstrating diffraction of microwaves. The intensity of the diffracted wave energy is indicated by the size of the current in the milliammeter.

Radio waves are diffracted around hills if these are separated by distances of the order of magnitude of the wavelength of the radio waves.

Radio waves can also interfere when a ground wave meets another wave that has taken a different route (figure 19.7), perhaps

ionosphere

The ionosphere is a layer of ions found in the upper atmosphere. It is brought about by the ionisation of gas molecules by energetic UV photons streaming down from the Sun. It can reflect radio waves of frequency less than 30 MHz.

travelling upwards and then being reflected at the **ionosphere** (a 'sky' wave). Depending on the height of the reflecting layer from the ground, the result at the receiver may be either a strengthening (constructive interference) or a weakening (destructive interference) of the signal received.

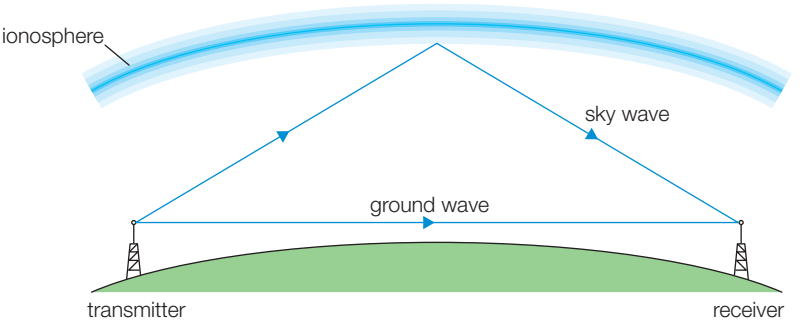


Figure 19.7 Interference of radio waves: a wave travelling near to the ground may interfere (constructively or destructively) with a wave reflected from the ionosphere.

ITQ6

What is the minimum wavelength of the radio waves which are reflected by the ionosphere?
Use the internet or look at an old radio to find out to which of the bands FM, AM, MW and LW these waves belong.

Because of the very small size of their wavelengths, it is difficult to observe effects such as interference and diffraction for waves in the gamma-ray spectrum. However, these rays can be identified as belonging to the electromagnetic family from the fact that their speed in air can be measured and has been found to be that of light. This is sufficient evidence that gamma rays are electromagnetic: no other type of wave has this characteristic speed in air!

Any wave whose speed in air is $3.0 \times 10^8\,m\,s^{-1}$ is an electromagnetic wave.

ITQ7

Do you know how inter-atomic spacing was first determined and which scientist was among the first to make this determination?

Chapter summary

- The Sun emits a wide range of electromagnetic waves because of its very high temperature.
- The energy transmitted by electromagnetic waves is called electromagnetic radiation.
- Electromagnetic waves are produced when electric charges are accelerated or decelerated.
- Electromagnetic waves transmit energy that is partly electrical in nature and partly magnetic.
- The oscillations of an electrical quantity and a magnetic quantity are in phase at all times.
- The oscillations of the electrical quantity and the magnetic quantity are both transverse.
- The direction of travel of an electromagnetic wave is perpendicular to the planes containing the two oscillations.
- Electromagnetic waves can travel in empty space. They do not need a medium for propagation.
- The speed of all forms of electromagnetic waves in a vacuum (or free space) is $3.0 \times 10^8\,m\,s^{-1}$.
- The Sun emits a wide range of electromagnetic waves.
- The energy transmitted by electromagnetic waves is called electromagnetic radiation.
- In other media in which the wave will travel, the speed is less than this value.
- The complete spectrum of electromagnetic waves ranges from a wavelength in air of about $10^4\,m$, falling continuously to about $10^{-15}\,m$. This corresponds to a range of frequencies from about $10^4\,Hz$ up to about $10^{23}\,Hz$.

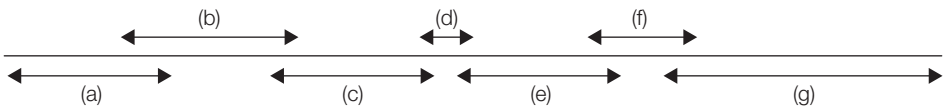
- The full spectrum may be divided up into smaller spectra, each one of these consisting of waves that all have the same effect on matter.
- These smaller spectra overlap considerably.
- Electromagnetic waves can be reflected, refracted, diffracted and made to interfere, in appropriate conditions.
- All the various forms of electromagnetic radiation in the full spectrum, except gamma rays, may be made to show the above-mentioned effects. The wavelengths of gamma rays are much too small to create the conditions required to show these effects.

Answers to ITQs

ITQ1 1.0×10^{-6} m
ITQ2 About 6000°C
ITQ3 3 m
ITQ4 Yes, since the frequency = 0.6×10^{12} or 0.6 THz
ITQ5 For the radio programmes to reach consumers the radio waves must reach them. The hills will not necessarily 'obstruct' the waves, but cause them to be diffracted. This will most easily happen if λ is about the same size as d, that is, if the wavelength of the radio waves is about the same size as the spaces between the hills. The spaces could be very large and so the wavelength should also be very large. Hence long-wave radio is required.
ITQ6 Minimum wavelength = 10 m. FM.
ITQ7 By taking photographs of diffraction caused by atoms. Max von Laue

Examination-style questions

1 The diagram shows the bands (a) to (g) of waves that make up the electromagnetic spectrum. Waves (a) have the shortest wavelengths in the spectrum and waves (g) the longest.



(i) Explain the meaning of the term 'spectrum'.

(ii) Why are the waves described as 'electromagnetic'?

In the diagram (b) denotes X-rays.

(iii) What do (c), (d) and (e) each represent?

The speed of all these waves in a vacuum is 3.0×10^8 m s^{-1}.

(iv) Calculate the frequency of a wave in band (g) whose wavelength is 600 m.

(v) Indicate by an arrow on the diagram roughly where this wavelength might lie along the spectrum.

(vi) How might waves in band (d) be detected?

(vii) How might waves in band (e) be produced?

(viii) For which band of waves can an opening of about 6 cm produce diffraction in a parallel beam?

2 Copy the table below and complete it by inserting a tick (✓) if the statement is true and a cross (×) if the statement is false, for sound and light waves.

	Sound waves	Light waves
Are transverse waves		
Can travel in a vacuum		
Energy is carried by particles only		
Can be absorbed by matter		
Can undergo refraction		
Can show interference		
Obey the laws of reflection and refraction		

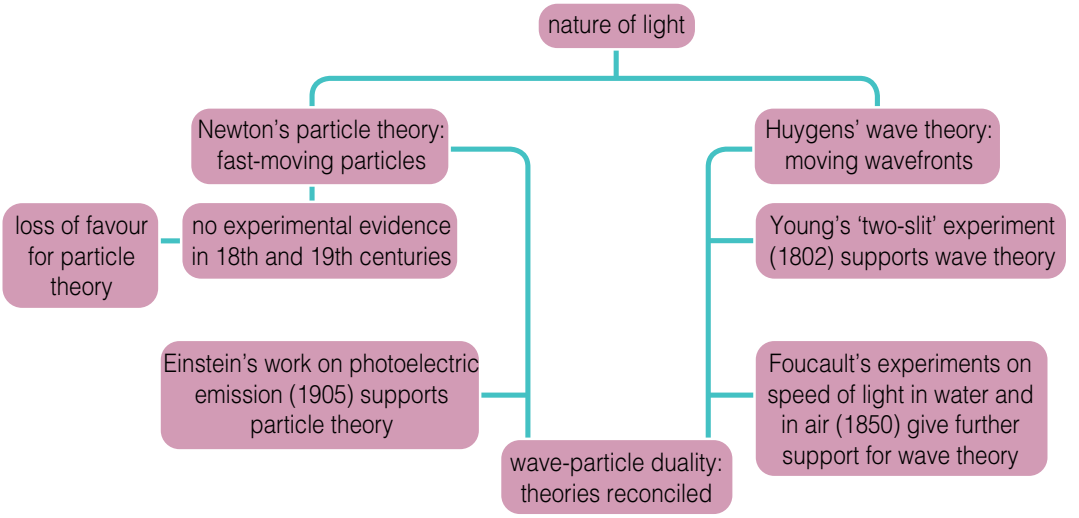
3 Is it possible to generate electromagnetic waves by charging a plastic comb by friction and waving it to and fro two or three times a second in the air? Explain your answer.

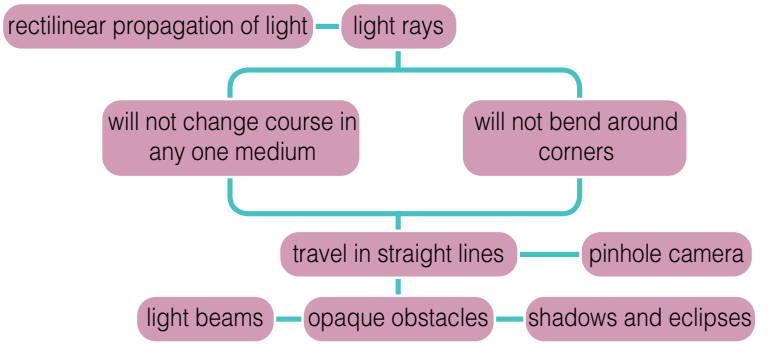
4 The wavelengths of light waves are normally measured in nanometres (nm), where 1 nm $= 1.0 \times 10^{-9}$ m. Assuming that one wavelength in green light is 550 nm, calculate the frequency of the light of that wavelength.

20 Light

By the end of this chapter you should be able to:

- recall the rival theories of light and compare them
- explain why the diffraction of light is not normally observed
- describe a simple 'Young's slits' experiment to show that light is a wave motion
- explain the significance of the results of this experiment in supporting the wave theory of light
- recall the significance of photoelectric emission in supporting the particle theory of light
- appreciate that light may be considered to behave like both waves and particles, the particles being packets or pulses of wave energy
- recall that each of these packets is called a 'photon' of light energy
- explain what is meant by the rectilinear propagation of light
- appreciate that the fact that light travels in straight lines is responsible for many everyday observations, including the formation of shadows and eclipses
- explain the formation of shadows and eclipses
- distinguish between a solar eclipse and a lunar eclipse
- describe how a pinhole camera works





Introduction

We currently acknowledge that light is a wave motion and that it is a form of electromagnetic energy. However, right up to the beginning of the 19th century, and in support of Newton's view, scientists thought that light consisted of very tiny particles that left an object at high speed and entered an observer's eyes, thus enabling him or her to see. This theory went out of favour when Thomas Young showed in 1802 that light could interfere, giving darkness in certain places and increased brightness in others. It was difficult to see how particles arriving at the same place could cancel each other out. We now think of light as having a 'dual nature', behaving like waves in some circumstances and like particles in others. The 'particles' we now take to be discrete packets or pulses of wave energy, called **photons**, which travel at the speed of light. No longer do we argue about the nature of light. We accept that there are circumstances in which light shows a wave nature and there are circumstances in which light behaves as though it was a stream of particles, these particles we now recognise as photons.

photons ›

Rival theories of light – a touch of history

In the latter half of the 18th century there were two rival theories about the nature of light namely, Isaac Newton's particle theory, and the wave theory of Christian Huygens, a famous Dutch scientist of the period. Most scientists agreed with Newton's idea, described in his publication Opticks (1704), that light consisted of minute particles travelling at very high speed from the object being observed to the observer's eye, thus enabling him or her to see. Newton explained reflection and refraction using this concept, and for more than a century the theory remained unchallenged, mainly because of Newton's authority as mathematician and scientist.

Towards the end of the 17th century, Huygens advanced a theory that light could be considered to be a wave formation in which each point on a wavefront (see pages 257 and 258) was a source of 'secondary waves'. This idea came to be known as Huygens' Principle, and is still used today to explain interference and diffraction.

In 1802 an English medical doctor named Thomas Young carried out an experiment in which he allowed light from two very small pinholes to interfere on a screen. He found that there were places on the screen where the light beams seemed to strengthen each other; and there were other places where the beams seemed to cancel each other out to give darkness. This effect could only be explained, he argued, if the light consisted of waves; whereas two waves could cancel each other out, two particles of light could not. This was the first evidence in support of the wave theory.

In spite of this, however, the particle theory continued to hold sway. It was not until 1850, when a French scientist, Jean Foucault, showed that light travelled faster in air than it did in water that the argument for the wave theory seemed to be clinched. Newton had predicted that light would travel faster in water, and this was shown to be false, and so his theory began to lose support. For a long time it seemed that the controversy was settled: light was, after all, a wave motion, as Huygens had maintained.

It was shortly after the beginning of the 20th century that people realised that both theories could be correct, as a result of the work of another scientific genius, Albert Einstein.

Current theory of light: wave–particle duality

Light shows some behaviour that can be easily explained if we assume that it consists of waves. For example, interference and diffraction are both explained by assuming that wave-like vibrations, when superposed, may result in an enhanced or a reduced amplitude depending on the phase relationship of the vibrations (see chapter 18). A particle theory of light could not easily explain such behaviour. But there are other phenomena that can be explained only by assuming that light consists of particles. One such phenomenon is the 'photoelectric effect' investigated by Albert Einstein in 1905.

photoelectric effect ›

Photoelectric emission, or the **photoelectric effect**, is the release of electrons from the surface of a metal when electromagnetic energy of a certain minimum frequency is shone on to it. It was found that no electrons were emitted from a given surface if the frequency of the radiation was any lower than this minimum, no matter how intense the beam of radiation.

To explain this, Einstein used an idea suggested earlier by another German scientist, Max Planck, that electromagnetic energy was 'quantised', that is, delivered in small packets, called '**quanta**', and not as a continuous stream. The amount of energy in a '**quantum**' (the singular of 'quanta') was related to the frequency of the electromagnetic radiation. Einstein proposed that an electron was emitted only if the energy packets or quanta associated with the electromagnetic radiation had sufficient energy to release an electron from the metal. Thus, if the frequency of the electromagnetic energy was too low, no electrons would be emitted, no matter how intense the beam.

quanta ›
quantum ›

The picture we now have of light is a stream of wave pulses leaving the light source in very quick succession. Each pulse takes with it an amount of energy that depends only on the frequency of the light. These pulses or packets of electromagnetic energy are called photons or quanta. They travel away from the light source at a speed of $3.0 \times 10^8\,m\,s^{-1}$ in air and in straight lines in a given medium.

So, we may think of light as being made up of both waves and particles at the same time. Each pulse of waves carries a 'particle' of energy (and not a particle of matter, as Newton and other earlier scientists would have thought). Whether the wave nature or the particle nature is more evident depends on the circumstances. In photoelectric emission, the particle nature dominates; and the wave nature dominates when the conditions are suitable for interference and diffraction to occur. We call this 'double nature' that light displays the **wave–particle duality** of light, and we can regard it as a reconciliation of the two rival theories of light. The particle theory of Newton and the early scientists and the wave theory of Huygens and Young can co-exist, after all!

wave–particle duality ›

Use of the photon in modern technology

Now that the doubt about the nature of light has been removed and we know that there are times and circumstances when one or the other of the two behaviours will predominate, much use has been made of the 'photon concept' in modern technology. One example of this is in the digital camera.

Ever since Einstein was able to show that the energy of photons could result in electron release, much use has been made of the principle. The idea is now used in the digital camera in which the arrival of focussed photons on a light sensor results in a flow of current which, with the help of computer technology, is able to produce a likeness (or a picture) of the subject (the person whose picture was being taken). The method of doing this is as follows.

You will know that in conventional cameras light from the subject is focussed on to photographic film by the camera lens and the film 'senses' the arrival of this light by chemical action. Chemical changes then begin in the film to produce the picture. In the digital camera there is still a lens focussing the light on to a 'sensor' (without a lens you cannot have a camera), but here the 'sensor' is not chemical, as in photographic film, but, 'physical' in the form of a 'photodiode'. The photo diode is a semiconducting device which conducts current when light falls on it rather than when a voltage is applied to it – hence its name 'photodiode', the prefix 'photo' suggesting the presence of light. This photodiode (the light sensor) has an array of thousands of 'photosites', each of which produces a very small current that, in turn, provides a small amount of image information (a 'pixel'). A digital picture consists of thousands of these pixels. These thousands of pixels together form the image.

The current resulting from the arrival of light photons on the photodiode depends only on the number of photons arriving each second on its surface and this, in turn, depends on the strength of the light leaving the subject. This resulting current is therefore a 'physical replica' of the subject, and not a chemical one, as in the conventional camera. This physical replica, the current, is now used by the computer to give an image of the subject which can be displayed on the screen as any text can, or printed, again as any text can.

Light behaving like a wave: Young's slits

When Thomas Young did his experiment, he did not use a double slit at all. Instead, he used two pinholes. However the use of two very narrow slits is much easier and the results are quite striking.

In 1802 Thomas Young performed what is now called the Young's slits experiment. We can do a similar experiment as follows (in a room that can be blacked-out).

1 Blacken one side of a glass slide by holding it in the smoke of a candle flame or by treating one surface of the slide with aquadag.

2 To make the pair of slits, hold two razor blades together with their edges parallel and pull them along the blackened surface of the glass slide. The slits will be very close together (less than 0.5 mm apart). Mount the slide in a retort clamp so that the slits are vertical.

3 Set up a straight filament lamp in another stand about 50 cm in front of the slits so that the filament is parallel to the slits.

4 Set up (also in a clamp) a sheet of red filter between the lamp and the slits.

5 Make the screen by taping a sheet of greaseproof paper over a stiff cardboard frame about 20 cm square. Support this screen in a clamp about 1 m behind the plane of the slits and parallel to it (figure 20.1 (a)).

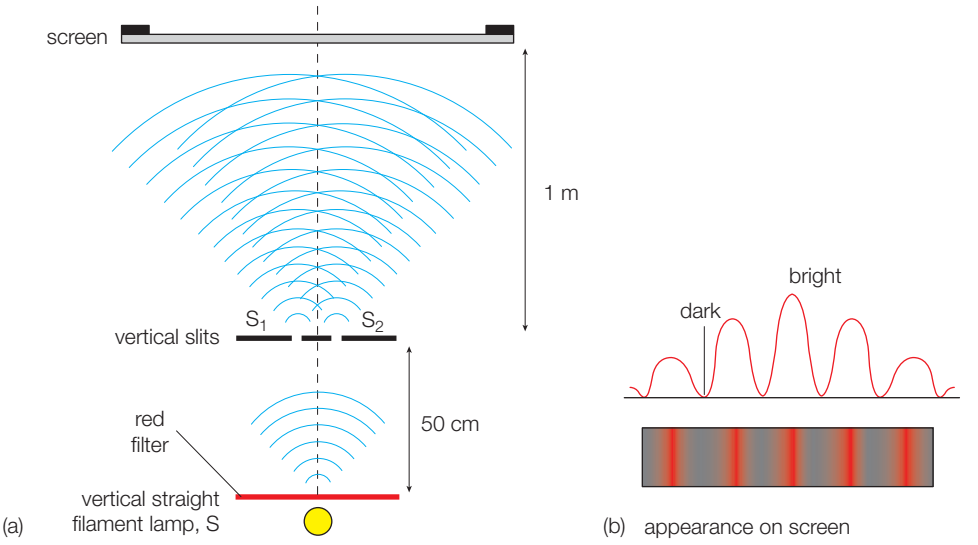


Figure 20.1 Young's slits experiment, not to scale. The slits are vertical as is the straight lamp filament. The fringes are, therefore, also vertical. (a) Top view of the set-up. (b) The fringes viewed from behind the screen.

6 Check to ensure that the line joining the source and the slits is perpendicular to the plane of the screen.

7 Switch on the straight filament lamp.

8 Use a high-power magnifier (a thick convex lens) to view the interference pattern from behind the screen. This pattern should be a series of red and dark bands (which Young called 'fringes', the term we still use today to describe the pattern). With proper adjustment of source, slits and screen, a set of vertical fringes should be seen (see figure 20.1 (b)).

This experiment is not as easy to do and get results from as the method just outlined might suggest. For best results the following 'precautionary' steps should be taken:

(i) The room should be thoroughly 'blacked out' (why is this necessary?).

(ii) Since the amount of light passing through the slits will be very small, it would help tremendously if the light source had a high 'wattage'.

(iii) The distance between the source and the slits should not be too large, (why?)

(iv) The distance between the slits and the screen should be reasonably large (for larger fringes), always remembering that the larger this distance, the less distinct will be the fringes.

(v) Most importantly, ensure that the straight filament and the slits are parallel and that the plane of the slits is parallel to the screen.

Alternating red and dark bands will be seen on the screen when viewed from behind with the magnifying glass. As already indicated, the separation of the fringes will depend on the distance between the slits and the screen.

Figure 20.1 (b) shows the appearance of the observed pattern and how it can be explained in terms of the wave nature of light. Each slit acts as a secondary wave source (see figure 18.18 on page 277) and the concentric arcs shown between the filament and the slits are wavefronts advancing towards the slits from the slits. These wavefronts reach the slits and the points where the meeting takes place behave like 'secondary sources' (Huygens' principle) and themselves send out secondary waves towards the screen. On the way to the screen the two sets of wavefronts meet and at every point of meeting, the waves will be in phase and produce constructive interference. If, therefore, the screen is placed parallel to the plane of the slits, there must be points on it where such meeting of wavefronts occurs all the time. This should happen for any position of the screen behind the pair of slits. These are the points where there will be bright red bands. These are the points of constructive interference.

Halfway between points of constructive interference on the screen there will be points of destructive interference, since these will be the points where crest meets trough and there will be no resultant. At such points there will be dark fringes. Remember that darkness means absence of light. So since there will be no resultant amplitude when a crest meets a trough, there will be no light there and darkness will result. It will be clear from the diagram that the nearer the screen is to the slits the smaller will be the separation of the bright fringes and also that of the dark fringes. The separation of the fringes therefore depends on the distance between the plane of the slits and the screen.

The graph at extreme right shows how the brightness of the fringes changes across the fringe pattern. The height of the peaks is a measure of the brightness of the fringes at their centres. As one moves away from the centre of the pattern the brightness of the fringes falls.

Since light may be regarded as a wave motion, we should expect it to be diffracted at edges. But when we consider that the extent of diffraction depends importantly on the size of the wavelength of the wave, we can see why any diffraction that light might undergo at edges and on passing through pinholes and nail holes will be negligible. (Remember that the amount of diffraction depends on the ratio λ/d.)

Compare this with the behaviour of sound whose wavelengths are millions of times larger than those of light.

Light travels in straight lines

rectilinear propagation

Photons leaving a light source travel in straight lines called rays. These rays are the paths or directions along which the energy flows. In this section we consider the behaviour of light as a 'straight-line' phenomenon, paying little attention to its wave nature. The movement of light along straight lines is sometimes called '**rectilinear propagation**'. 'Ray optics' is the study of the behaviour of light assuming that the light travels in straight lines.

Evidence that light travels in straight lines is found in the formation of shadows and eclipses, the formation of images by pinhole cameras and the fact that we cannot see around corners. The first simple experiment to show that light travels in straight lines may be carried out as follows:

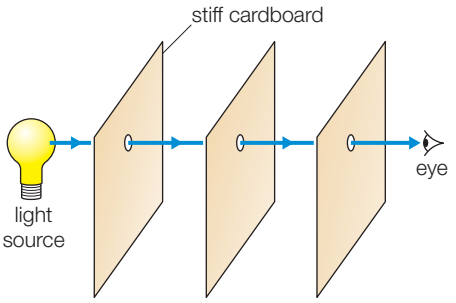


Figure 20.2 Demonstrating that light travels in straight lines.

1 Make a tiny hole in each of three similar sheets of stiff cardboard, the same distance from one of the short edges.
2 Support each sheet in the clamp of a retort stand with three holes roughly in line (figure 20.2).
3 Pass a length of string through each of the holes and pull the string taut. The holes will now be in line.
4 Now clamp the cardboard sheets firmly in position and remove the length of string.
5 Place a lighted lamp opposite the hole in one of the outer sheets.
6 Look through the hole in the other outer sheet towards the lamp. You should see the lighted lamp.
7 Now move the middle sheet very slightly out of line. The lamp can no longer be seen.
8 The lamp can be seen only when the three holes are in line.

We conclude that light has travelled in a straight line in the experiment.

Convergent and divergent beams

light beam or bundle

divergent

convergent

A set of light rays travelling together is called a **light beam** or (sometimes) a light bundle. If the rays are parallel, we say the beam is a parallel beam. If the rays are spreading out, the beam is **divergent**; and if they are coming together, the beam is **convergent** (figure 20.3).

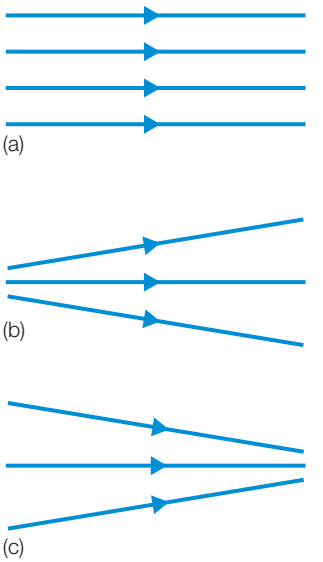


Figure 20.3 Types of light beam: (a) parallel; (b) divergent; (c) convergent.

Shadows

Because light travels in straight lines, a shadow is formed when an opaque object is placed in the path of a beam of light rays; the light rays cannot bend around the object.

The type of shadow formed depends on whether the light comes from a point source or an extended source (figure 20.4).

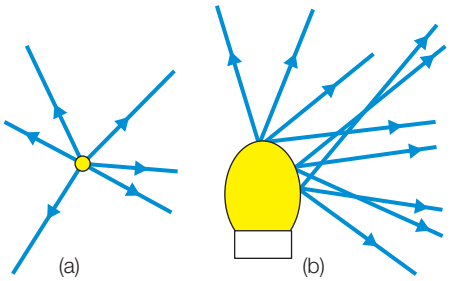


Figure 20.4 (a) A point source of light. (b) An extended source of light.

Shadow formed when a point source is used.

point source

A **point source** is one that is so small that we can assume that all the rays of light come from the same point (figure 20.4 (a)).

The shadow formed on the white screen in figure 20.5 by placing an opaque object in the path of rays from a point source is of uniform darkness, since no

umbra ›

light whatever from the source falls on the area covered by the shadow. Such a shadow is called an **umbra**.

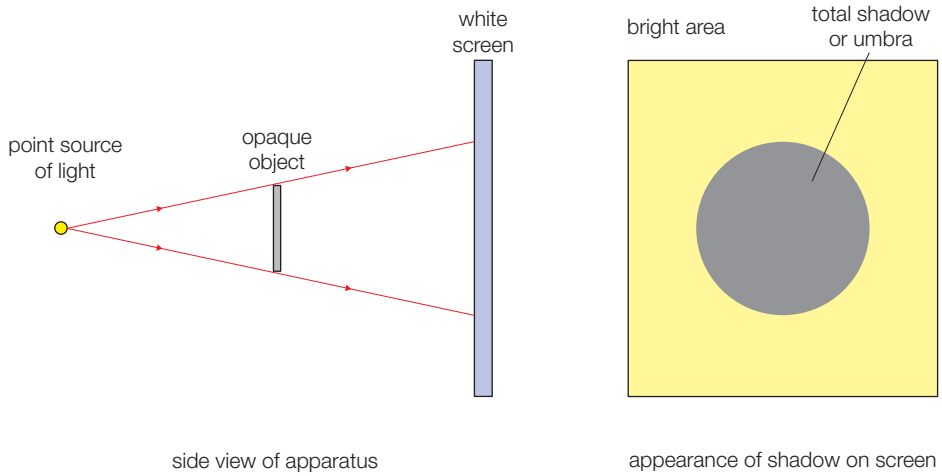


Figure 20.5 Shadow formed by a point source of light.

Shadow formed when an extended source of light is used

extended source ›

An **extended source** is sufficiently large for rays to arrive at an object from different points on the source. The lamp shown in figure 20.4 (b) may act as an extended source.

If the same opaque object as before is placed in the path of light from an extended source (figure 20.6), two types of shadow are seen on the screen: a total shadow or umbra of uniform darkness in the centre, and a partial shadow,

penumbra ›

called a **penumbra**, around the umbra. No light form the source reaches the screen at any point within the umbra, and light from some parts of the extended source does fall on the screen in the penumbra. The further one goes outwards from the umbra, the greater the amount of light falling on the screen, and therefore the lighter the penumbra becomes.

Umbra is the Latin word for shade. 'Penumbra' is derived from the Latin word *paene*, meaning 'almost', and *umbra* (shadow).

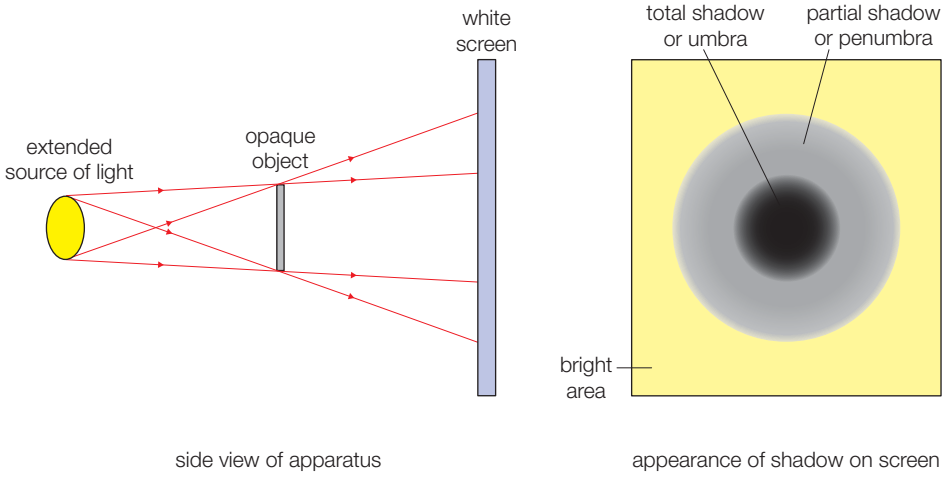


Figure 20.6 Shadow formed by an extended source of light.

There is really no sharp dividing line between the umbra and the penumbra, nor is there a sharp dividing line between the penumbra and the bright surround. In fact, the degree of darkness slowly gets less and less as one moves outwards from the umbra. You can check this by looking carefully at the shadows formed by large street signs of large objects. Figure 20.6 does not quite show this.

Eclipses of the Sun and the Moon

IT Q1

Look at figure 20.7 (b). What sort of solar eclipse will be seen by people on Earth: (i) in region A; (ii) in region B? (iii) Would someone in region C observe an eclipse?

Eclipses of the Sun and the Moon are produced by the shadows cast by the Moon and the Earth, respectively. They all take place as a result of the fact that light travels in straight lines.

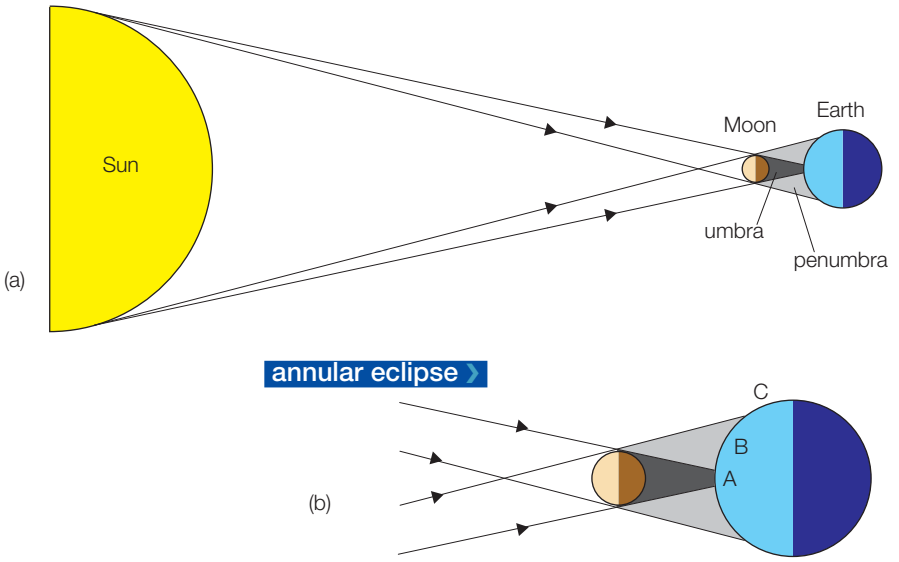
Because the Moon is so much nearer the Earth than the Sun, they both look roughly the same size from the Earth. We say that they 'subtend the same angle' at the Earth. In reality the Sun is vastly larger than the Moon.

Solar eclipse

If the Moon passes between the Earth and the Sun, the shadow of the Moon moves across the Earth, blocking out the light from the Sun. We then have a **solar eclipse**. The Sun is an extended source, so some parts of the Earth are in the Sun's umbra and some other parts are in the penumbra (figure 20.7 (a)). People in the umbra on the Earth (figure 20.7 (b)) see a total eclipse, because there is no light where they are situated; those in the penumbra see a partial eclipse, since they receive some light from the Sun – how much they get depends on where they are. People in other regions do not see any eclipse at all.

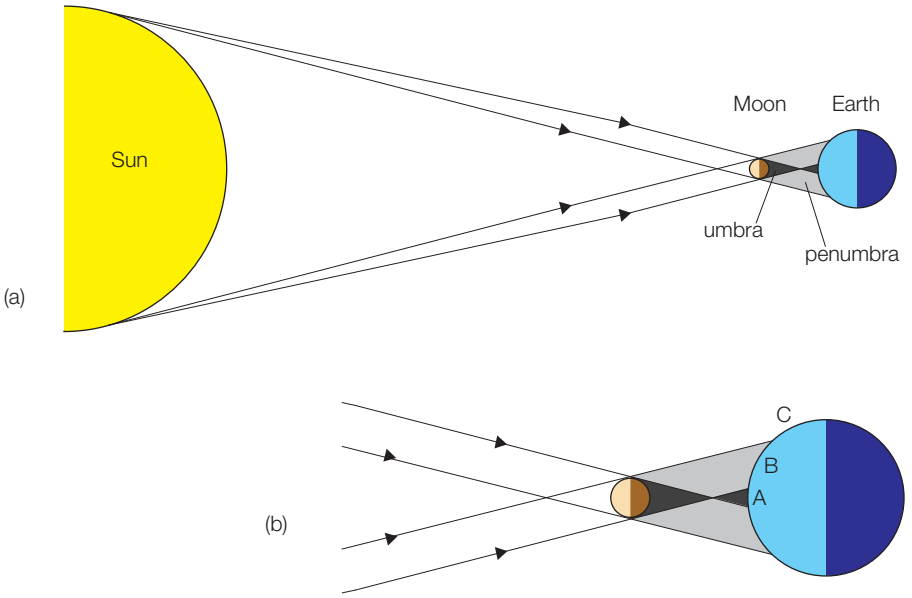
Sometimes the Moon is too far away from the Earth to produce a complete eclipse, and we see instead a dark disc surrounded by a bright ring. This is called an **annular eclipse** (figure 20.8), since what is seen is a bright ring around the Moon, and the Latin word for ring is 'annulus'.

solar eclipse ›



annular eclipse ›

Figure 20.7 Eclipse of the Sun.




(c)

Figure 20.8 (a) and (b) Annular eclipse of the Sun. (c) Annular eclipse as seen by people in region A on Earth.

Lunar eclipse

A lunar eclipse (or eclipse of the Moon) occurs when the Earth is between the Sun and the Moon and the three are in line (figure 20.9). As the Moon orbits into the Earth's penumbra, it becomes less bright; and when it enters the umbra completely, it is fully eclipsed and should disappear from view but often appears as a red disc . It then leaves the umbra and re-enters the penumbra on the other side, getting brighter and brighter again as it moves out of the penumbra, since the penumbra gets increasingly bright as its outer edge is approached.

Even during a total lunar eclipse it is still just possible to see the Moon. This is because some light from the Sun is refracted by the Earth's atmosphere towards the Moon. The Moon then can appear to be a dark-red or rusty colour at the time of total eclipse.

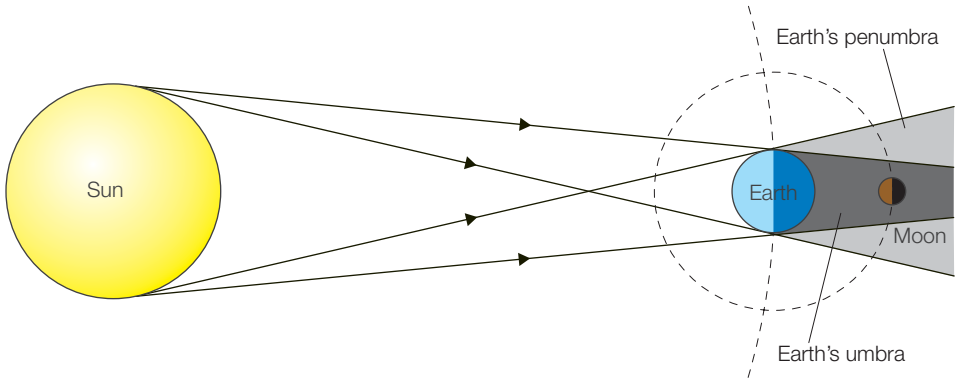


Figure 20.9 An eclipse of the Moon.

A partial lunar eclipse does not mean that a part of the Moon is blocked from view from Earth as happens with the Sun during a partial solar eclipse. 'Partial' here means that the Moon is not as bright as it normally is, because light is reaching it from only part of the face of the Sun' that is, the Moon is partially shadowed by the Earth. In other words, the Moon is in partial eclipse when it is in the Earth's penumbra.

Because of the Moon's small size compared with that of the Earth, it can take quite a long time to pass through the umbra, often much longer than an hour.

Because of the relative positions of the Sun, Earth and Moon at the time of lunar eclipse, the only moon phase at which an eclipse is possible is the full moon. However, an eclipse does not happen at every full moon because the three bodies are not exactly in line every time there is a full moon. The Moon's orbit round the Earth is usually somewhat tilted with respect to the Earth's orbit round the Sun.

The pinhole camera

A simple pinhole camera is a convenient way of demonstrating that light travels in straight lines (rectilinear propagation).

Preparation of the camera

The inside surfaces of a sturdy cardboard box are lined with black paper. A small circular hole of about 5 cm diameter is cut in the front of the box and a piece of thin, stiff cardboard or black paper is pasted over the hole (figure 20.10). A tiny hole is then made in the centre of this card or black paper with a sewing needle. The cardboard at the back of the box is cut away as shown in figure 20.10, and a sheet of crumple-free white tracing paper or greaseproof paper is pasted over the opening to serve as a screen. The box is carefully sealed with tape to make it lightproof.

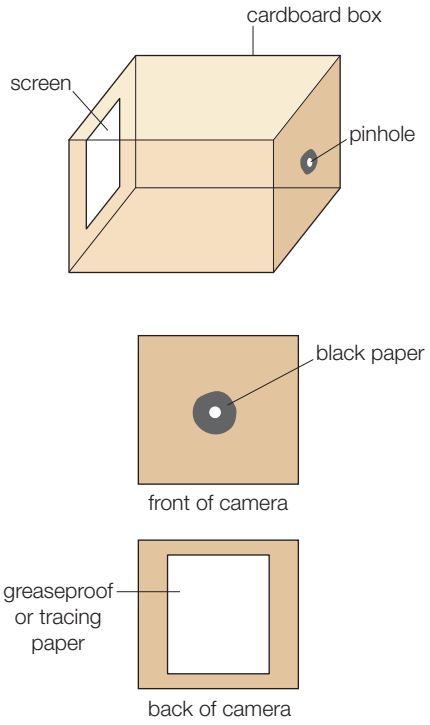


Figure 20.10 A pinhole camera.

Obtaining an image

An object is chosen that is self-luminous or, otherwise, is well illuminated from in front with a powerful source of light such as a 60 W electric light bulb.

Suppose the object in figure 20.11 is a small illuminated 12 V motor car headlamp bulb with a straight filament. A clear and inverted image of the filament is seen on the screen of the pinhole camera. The room need not be darkened, although the darker it is, the more clearly will the image be seen.

real image ›

An image that is formed on a screen is called a **real image** since, in order for that image to appear, light rays must have fallen on the screen.

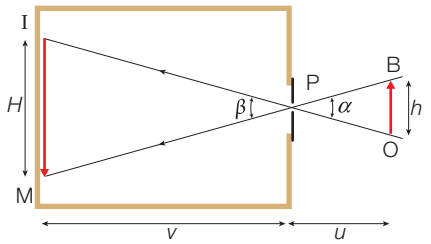


Figure 20.11 Formation of an image by a pinhole camera.

The lamp bulb is placed in such a position in front of the camera that the image formed on the screen is large enough to be clearly made out. Ensure that the object (the lamp filament) is directly in front of the pinhole and position it near enough to the pinhole to obtain an image that occupies almost the whole height of the screen

Investigating features of the image

Size

The size of the image can be changed by moving the object away from the camera or towards it. You will find that moving the object away from the camera reduces the size of the image.

Magnification

This is the factor by which the image obtained is larger than the object.

When you have a satisfactory image:

1 Measure the size of the lamp filament (h).
2 Measure the size of the image of the filament on the screen (H).
3 Measure the distance from the object to the pinhole (u).
4 Measure the distance from the image to the pinhole (v).
5 Calculate the ratio H/h and the ratio v/u. Compare these ratios.

ITQ2

Why is it a good idea to paint the inside of the pinhole camera matt black?

You should find that the two ratios are equal, allowing for errors of measurement and judgment as to when you have the best defined image. It is also not easy to measure the size of the lamp filament, since you will have no access to it. (What sort of measurement error is very likely here?).

MATHEMATICS: similar triangles

If these two ratios, H/h and v/u are equal within the limits of error, it means that the two triangles POB and PIM of figure 20.11 are similar and the angles α and β, which are vertically opposite angles, are equal. This in turn implies that BPM and OPI are straight lines.

Orientation of the image with respect to the object

The ray diagram of figure 20.11 suggests that the image should be inverted about a horizontal axis. Careful examination of the image should show that this is so. If there is inversion about a horizontal axis there should also be inversion about a vertical axis.

We conclude that the rays from O and B passed through the pinhole without bending: they travelled through the pinhole along straight lines.

This demonstration shows that:

1 the light rays producing the image travel in straight lines;
2 where the rays from a point on the object meet the screen there is an image of that point;
3 the camera produces an image which, when seen from behind the camera, is inverted about the vertical and horizontal axes; and
4 the image produced by the camera is real, that is to say, it can be formed on a screen.

We could investigate what happens to the image when:

- the distance of the object from the camera is increased;
- the size of the pinhole is increased.

We would find that, as the object's distance from the camera is increased, the size of the image decreases. A ray diagram shows this clearly.

Effect of the size of the pinhole on the image

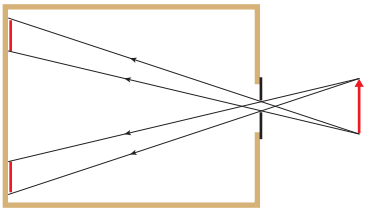


Figure 20.12 Formation of an image by a pinhole camera with an enlarged pinhole.

Increasing the size of the pinhole would make the image less sharp (i.e. it would become blurred), but brighter. This is because divergent rays, rather than a narrow nearly parallel beam, would leave each point on the object, as shown in figure 20.12. Where each of these rays meets the screen there would be an image of the point where the rays originated on the object. If the pinhole was very small, fewer rays will enter the camera and there would be less divergence (spreading out) between them; fewer and less diffuse images will be produced of the point on the object. This would make for better definition (sharpness) of the image. With a larger pinhole, more rays from a particular point would enter the camera and there would be more divergence between the rays; the image would then become blurred though, at the same time, it would be brighter, since more light rays would be admitted by a larger hole.

Fine pinholes produce well-defined images, but the drawback is that there is less light falling on the screen (fewer photons) than if the hole were bigger, so the image is dimmer.

If we needed a permanent picture of the object, we would use a photographic film and carry out the exposure for a long time (possibly hours). In such a case we would have to ensure that the camera did not move since, if it did while taking the picture, the image would be blurred.

ITQ3

The demonstration carried out with the pinhole camera showed that the image obtained was a sharp 'geometrical' image, that is to say, its form suggested that rays passing through the pinhole went straight through without bending. Why was the light not diffracted as it passed through the pinhole?

Chapter summary

- Up to the middle of the 19th century there were two rival theories of light – the particle theory and the wave theory.
- The particle or corpuscular theory, proposed by Isaac Newton, stated that light consisted of streams of very tiny fast-moving particles.
- Christiaan Huygens, a Dutch scientist, had proposed the theory that light was a wave motion.
- For a long time there was no evidence to support either theory, but support for the wave theory was first provided by Thomas Young's interference experiment in 1802.
- Further evidence for the wave theory was provided by Foucault, who showed that light travelled faster in air than in water, contrary to the prediction of Newton's particle theory.
- A very long time after Young's experiment supporting the wave theory, in 1905 Einstein's explanation of photoelectric emission provided the first piece of evidence that light could consist of particles.
- The current belief is that light can show both particle behaviour and wave behaviour depending on the circumstances.
- This behaviour of light ether as a particle or as a wave is called 'wave-particle duality' of light.
- The 'particles' of light are not particles of matter but packets of energy called photons.
- A ray of light may be thought of as the path along which a stream of photons travels.
- The movement of light particles (or photons) along a straight path is described as 'rectilinear propagation'.
- Shadows are formed when light is 'blocked off' by opaque objects.

- A shadow may be:
  - a total shadow, called an umbra, in which no light is present, or
  - partial shadow, called a penumbra, in which some light is present.
- An opaque object placed in the path of light from a point source produces only an umbra.
- An opaque object placed in the path of light from an extended source produces an umbra and a penumbra.
- Eclipses of the Sun are produced when light from the Sun is blocked off by the Moon.
- Eclipses of the Moon are produced when light from the Sun that would normally reach the Moon is blocked off by the Earth.
- The pinhole camera is the simplest device for obtaining an image of an object. It uses the principle of rectilinear propagation.

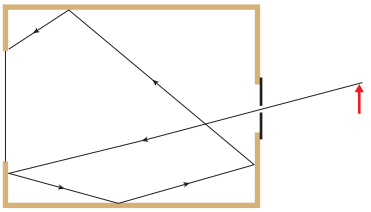


Figure 20.13

Answers to ITQs

ITQ1 **(i)** Total; **(ii)** partial; **(iii)** no.

ITQ2 To improve the quality of the image by ensuring that any incident light that is reflected from the screen does not reach it again through multiple reflection (figure 20.13). This would cause images of the same point to appear at different places on the screen instead of at one place only.

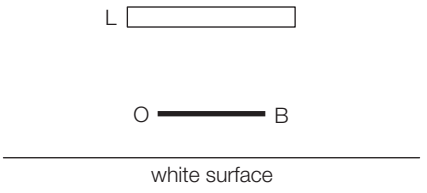
ITQ3 The diameter of the hole is probably many hundreds of times larger than the wavelength of the light passing through it. So λ/d is extremely small, and so any diffraction there might be is negligible.

Examination-style questions

1 Usually, light travels in straight lines. Under what conditions will light *not* travel in straight lines?

2 What do you understand by the term 'ray of light'?

3 The diagram on the left shows a lighted fluorescent tube, L, situated 3 m directly above an opaque object, OB. There is a white surface 1 m below the opaque object. Draw a labelled diagram to scale showing the umbra and the penumbra on the white surface. Show and label the width of each.



On your diagram show a point where the penumbra is:

- least dark; label the point X;
- darkest; label the point Y.

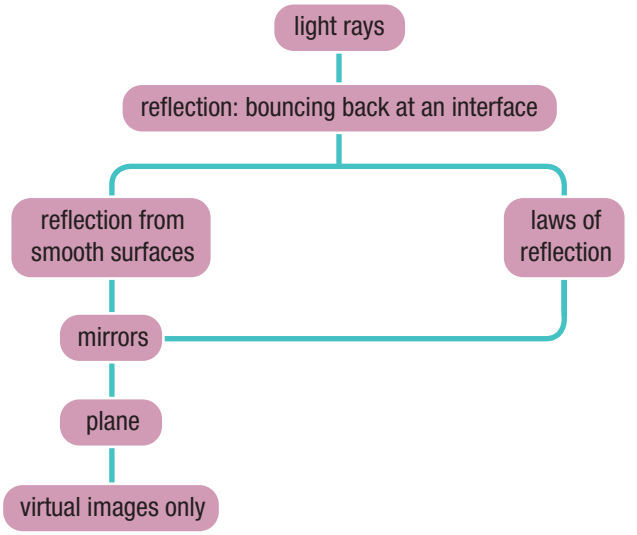
Use rays to show why Y is much darker than X.

4 With the aid of ray diagrams, explain why fluorescent lamps given softer, lighter shadows compared with those from incandescent bulbs, which give harsher, darker shadows.

21 Reflection of Light

By the end of this chapter you should be able to:

- state the laws that govern reflection
- perform experiments to test these laws
- recall what is meant by an image with regard to reflection
- use the laws of reflection to explain how images are formed
- distinguish between real and virtual images
- construct ray diagrams to show how virtual images are formed by plane mirrors
- find the position of a virtual image produced by a plane mirror by:
  - a ray tracing method
  - a no-parallax method
- recall the features of the virtual image produced by a plane mirror
- recall and use the fact that the distances of the object and image from a plane mirror are equal
- use the laws of reflection to solve problems



Introduction

We see most of the things around us by reflected light – that is, by light which bounces off surrounding surfaces. Most of this light, during the day, comes from the Sun. Although there are 'self-luminous' bodies, such as the Sun or the filament in a light bulb when it is switched on, we see most objects because light falls on them and is reflected off their surfaces into our eyes.

In this chapter we will consider how plane mirrors produce images, and the features of those images and where they are to be found.

The laws of reflection

Some of the terms we use when discussing reflection are illustrated in figure 21.1.

A ray of light strikes the mirror at the point P and leaves the mirror by reflection. The ray striking the mirror is called the **incident ray**. The point at which the ray strikes the mirror is the **point of incidence**, shown here as P.

The ray that bounces off the mirror is the **reflected ray**.

If a perpendicular to the mirror is drawn at the point P, that perpendicular is called the **normal** at P.

The angle (i) between the incident ray and the normal is the **angle of incidence**.

The angle (r) between the reflected ray and the normal is the **angle of reflection**.

The two laws governing reflection, called the laws of reflection, are:

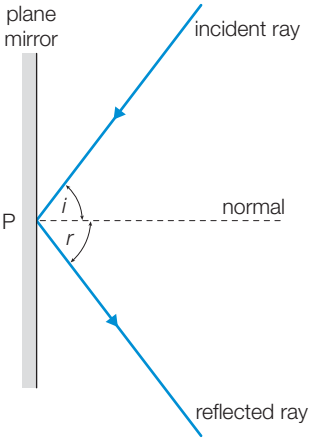


Figure 21.1 A ray of light is reflected at a flat (plane) mirror.

first law of reflection

1. **The incident ray, the reflected ray and the normal at the point of incidence all lie in the same plane. This plane is called the 'plane of incidence'.**

second law of reflection

2. **The angle of incidence is equal to the angle of reflection, or $i = r$**

To test the two laws of reflection, we can carry out Practical activity 21.1, which was first used in chapter 3 as a model for developing Measurement and Manipulation skills.

Practical activity 21.1

Verifying the laws of reflection

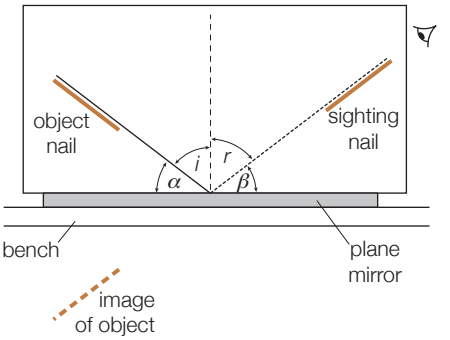


Figure 21.2 Verifying the laws of reflection. Side view of arrangement.

You will need:

- plane mirror
- two large headless nails about 6 cm long, or long pencils
- metre rule
- two retort stands with clamps
- large drawing board.

Method

1. Place the plane mirror flat on the laboratory bench with its reflecting surface uppermost.
2. Place a nail (the 'object' nail) with its head in the clamp of one of the stands and incline it over the plane mirror at an angle of about 45° to the mirror and about 3 cm away from it (figure 21.2).
3. Look at the image of the nail in the mirror and adjust your direction of view until you are looking along the image and can see only its tip.
4. Place the head of the other nail (the 'sighting' nail) in a clamp and adjust its position and orientation so that it is in line with the image of the first nail. When this is done, you should not see the image of the first nail at all as you look along the length of the sighting nail, since the second nail will be covering it.

5 Now carefully place the drawing board in such a position that both nails, still supported in the clamps, touch the board along their entire length.

6 Check that the plane of the board is perpendicular to the plane of the mirror.

7 Hold a length of string along the object nail so that one end of the string touches the mirror. Using a large protractor, measure the angle, α, the string makes with the mirror surface (see figure 21.3). Do the same with the sighting nail to measure angle β.

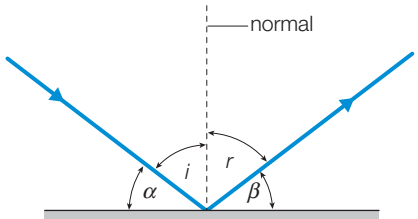
Results

When the sighting nail was positioned so that it blocked off the image of the first nail, then it was in line with the reflected ray. Both nails, representing the incident and the reflected rays, were in the plane of the board. If this plane was perpendicular to the mirror, then the normal to the mirror, the incident ray and the reflected ray were all in the same plane.

We can determine whether the values of α and β could be deemed to be equal by considering the limits of uncertainty of each angle. If the angles are found to be equal within these limits, then we must conclude that the angle of incidence is equal to the angle of reflection and that the second law is also true.

From figure 21.3 it can be seen that if the value of angle α is the same as that of angle β, then the value of the angle of incidence, *i*, is the same as that of the angle of reflection, *r*.

MATHEMATICS: complementary angles



If $\alpha = \beta$,
then $90° - i = 90° - r$,
and $i = r$

Figure 21.3 The angles α and β are called the 'glancing' angles.

Locating the image formed by a plane mirror

There are two ways of finding where the image formed by a plane mirror is situated:

- by ray tracing, and
- by using the method of 'no-parallax'.

In Practical activity 21.2 we will use ray tracing to find the position of the image. Practical activity 21.3 demonstrates the method of 'no-parallax'.

Practical activity 21.2

Finding the position of an image formed by a plane mirror by ray tracing

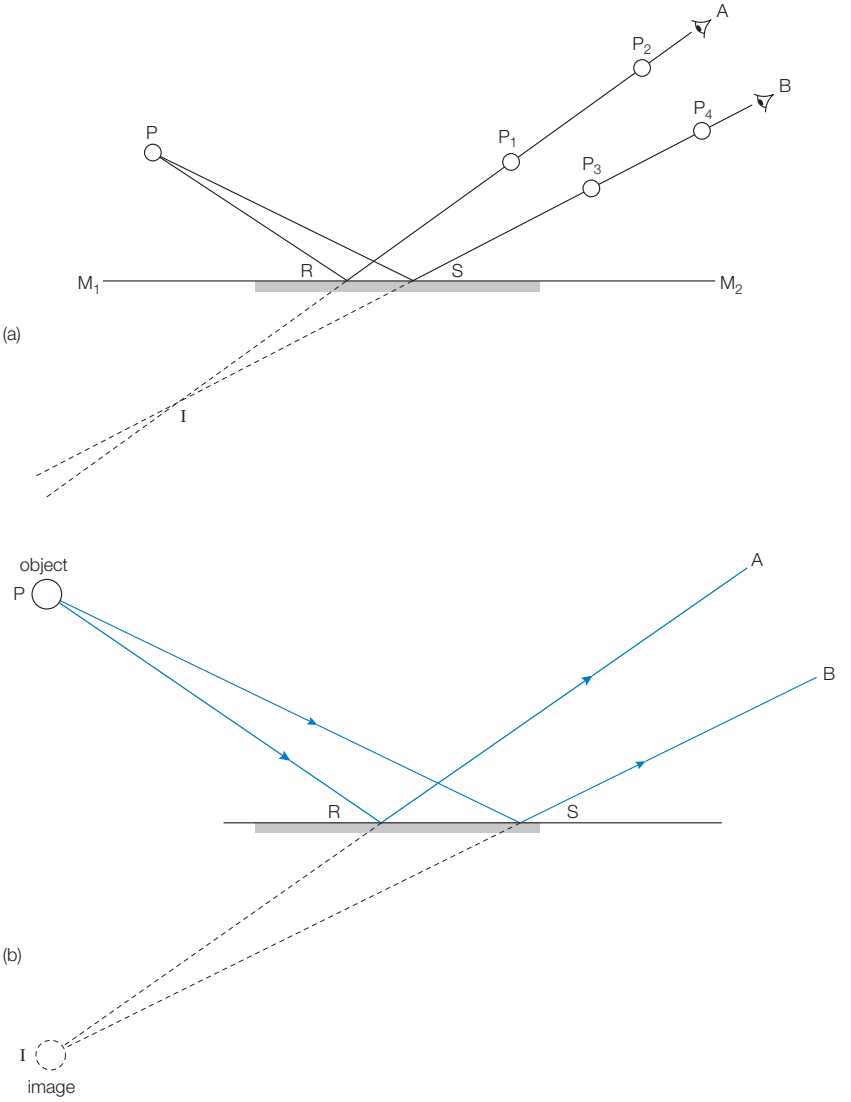


Figure 21.4 (a) Finding the position of the image produced by a plane mirror by ray tracing. (b) The incident and reflected rays. The image is a virtual one: the reflected rays only appear to come from I. No light from the object pin at P actually passes through I, and the image could not be formed on a screen. It is therefore a virtual image.

You will need:

- plane mirror (a metal mirror is preferable)
- adhesive tape
- sheet of paper
- Plasticine
- three optical pins
- large drawing board.

Method

1 Fix a sheet of paper to the drawing board on the bench top using adhesive tape.

2 Draw a line (M_1M_2 in figure 21.4 (a)) in the centre of the paper. Stand a plane mirror vertically on its edge along M_1M_2. Use Plasticine if necessary.

3 Stick an optical pin (the 'object' pin) vertically at some point P in front of the reflecting surface of the mirror and about 5 cm from it.

4 Stoop to get your eyes at the level of the mirror and, having closed one eye, look into the mirror with the other from direction A as shown in figure 21.4 (a). Stick two sighting pins into such positions, P_1 and P_2, that the two

sighting pins and the image of P all seem to be in line. For good accuracy the second sighting pin at P_2 should be fairly far away from the first sighting pin at P_1.

5 Draw small pencil circles round the pins at P_1 and P_2, and mark these points P_1 and P_2. Then remove the sighting pins from the paper.

6 Repeat steps 4 and 5, looking from a different direction (e.g. B in figure 21.4 (a)), and marking the new positions of the sighting pins by P_3 and P_4. Remove the sighting pins from the paper.

7 Draw a small pencil circle round the object pin at P, and mark the point P. Remove the object pin.

8 Draw a line joining the points P_1 and P_2 to meet the mirror line M_1M_2 and produce this line by a dashed line to a point well beyond the mirror line.

9 Do the same for the points P_3 and P_4 until the two dashed lines meet.

10 Mark the point of intersection of the dashed lines produced behind the mirror as I. This is the position of the image of the object at P.

Explanation

For the image to be seen, a ray that leaves the object must be reflected at the mirror and then enter the eye. The position of the image is somewhere along the direction of the reflected ray. Therefore, the image of P is somewhere along the line P_1P_2 and also along the line P_3P_4. The position of the image must, therefore, be at the intersection of these two diverging lines when they are produced backwards.

Practical activity 21.3

Finding the position of an image in a plane mirror by the no-parallax method

You will need:

- plane mirror (a metal mirror is preferred)
- adhesive tape
- sheet of paper
- Plasticine
- two optical pins
- large drawing board.

Method

1 Fix a sheet of paper to the drawing board on the bench top using adhesive tape.

2 Draw a line M_1M_2 in the centre of the paper. Stand a plane mirror vertically on its edge with its reflecting surface on the line M_1M_2 (see figure 21.5 (a)). Use Plasticine, if necessary, to keep the mirror steady.

3 Stick an optical pin (the 'locating' pin) vertically at some point P_1 behind the mirror and about 10 cm away from it.

4 Stoop so that your eye-level is that of the mirror and, with one eye closed, look towards the locating pin at P_1 in the direction perpendicular to the mirror (see figure 21.5 (a)). You should be able to see the top part of the pin at P_1, above the top edge of the mirror.

5 Place another pin (the 'object' pin), in front of the mirror in such a position that you can see the object pin (in front of you), the image of the object pin in the mirror and the top of the locating pin all in line to start with.

6 With one eye still closed, move your head sideways (parallel to the mirror) and look to see whether the image of the object pin seen in the mirror and the top part of the locating pin (seen above the mirror) move apart increasingly when you move your head from side to side. This will usually be the case (figure 21.5 (b)).

7 Repeat steps 5 and 6 with the object pin in different positions nearer to or further from the mirror, until you find a position where there is no parallax (which means 'no separation') between the image of the object pin and the locating pin P_1. The image stays in line with the locating pin, as shown in figure 21.5 (c), even when you move your head sideways.

8 Draw small circles in pencil round the locating pin and the object pin, and mark these points P_1 and P_0. Remove the pins and mirror.

9 Draw a line joining P_0 to P_1. Mark the point where this line intersects the mirror line as X.

10 Measure the lengths P_0X and P_1X.

Result

The distances P_0X and P_1X should be very nearly equal. In other words, the distance of the object from the mirror (the 'object distance') should be found to be equal to the distance of the image from the mirror (the 'image distance').

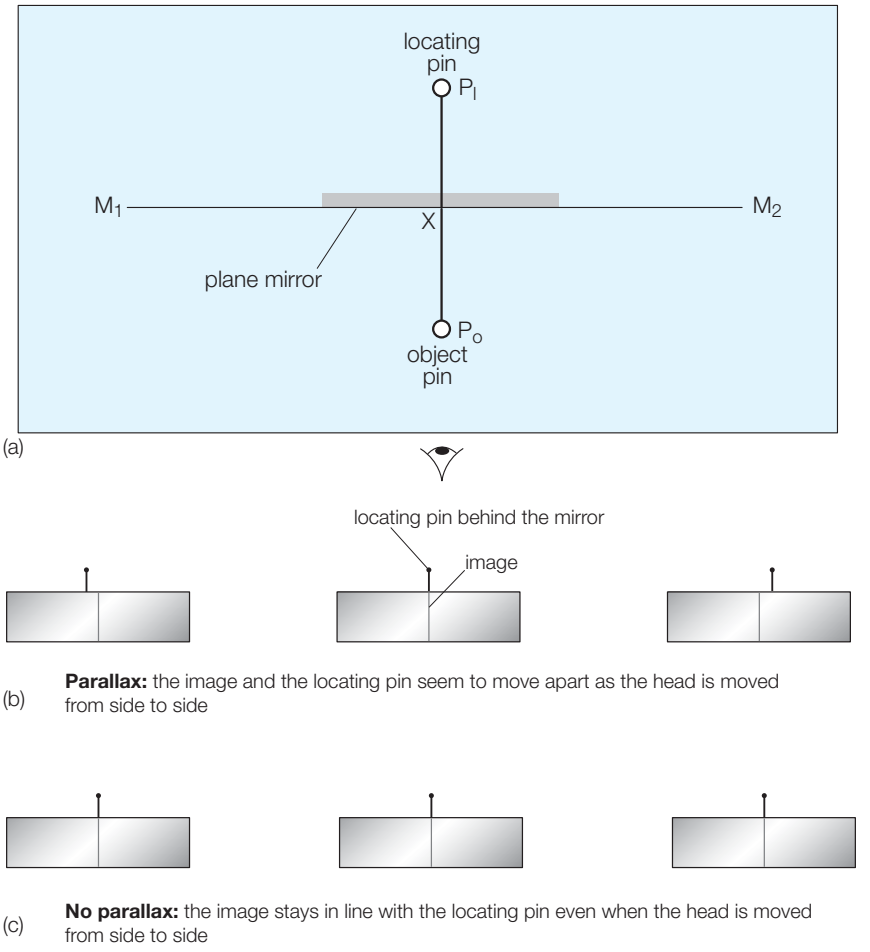


Figure 21.5 (a) Using the no-parallax method to find the position of the image formed by a plane mirror. View of arrangement from above. One eye is moved from side to side, observing the image of the object pin at P_0 until this image stays in line with the top of the locating pin at P_1. (b) Parallax is seen. (c) No parallax. The position of the image coincides with the position of the locating pin.

It is often the case, however, in such an experiment that P_oX, the object distance, and P_1X, the image distance, are not quite the same within the limits of measurement and judgment error. The reason for this could be that the effective reflecting surface of a glass mirror may not be the surface of the mirror, but a surface somewhere inside the mirror, this being due to the effect of refraction of the glass of the mirror. It is therefore better, in all experiments involving reflection, to use metal mirrors where a case of refraction will not arise.

Features of the image formed by a plane mirror

To understand the mechanism of formation of the image, look again at figure 21.4 (a). In that figure rays of light, PR and PS, leave a point P on the object and travel towards the mirror, M_1 M_2, then they strike the mirror at R and S, and are reflected from it along RA and SB respectively. Note that these

rays are diverging and will not meet anywhere. As long as the mirror acts on the rays an image will be formed, either real or virtual.

Where is the image formed and what is its nature?

The answers to these questions depend on whether the reflected rays are converging or diverging on reflection:

- **Case (i) –** if the rays converge on reflection (which is not the case here), they will meet somewhere and give a *real* image where they meet.
- **Case (ii) –** if the rays diverge, as is the case here, then they will not meet anywhere and will therefore give a *virtual* image at the point where they *seem* to have come from behind the mirror.

An image formed by an optical device is always either real or virtual, depending on how the rays behave on leaving the device.

In the case of figure 21.4 (a) it is clear that the rays diverge off the mirror and so the image will be virtual. The image is situated at the point where the reflected rays *seem* to have come from. To an eye receiving these rays it would be just the same as if there was something at I giving off the two diverging rays. Clearly, what the eye will see is at I, since this is the point where the rays will seem to have originated, and so this must be the position of the virtual image. The image is virtual because there are no rays behind the mirror and so a screen placed at I will show nothing. Such an image is said not to be real, since it cannot be shown on a screen. So the virtual image is really at I, the point where the dotted lines (which are construction lines and not rays!) meet behind the mirror. In drawing ray diagrams, construction lines are always dotted or dashed lines, never full lines. Only rays are represented by full lines.

The conclusion is, then, that the image of the object P is at I, the point where the reflected rays appear to have come from and it is virtual. This is always the procedure to be followed to find the position of an image formed by a plane mirror – produce backwards (by dashed lines) the rays leaving the device which has acted on those rays and the point at which the dashed lines meet will be the position of the virtual image.

To an eye positioned in the path of these reflected rays the image will appear to be 'in' the mirror, but, of course, will really be positioned at I.

Plane mirrors produce only one type of image, namely virtual, no matter where the object is placed.

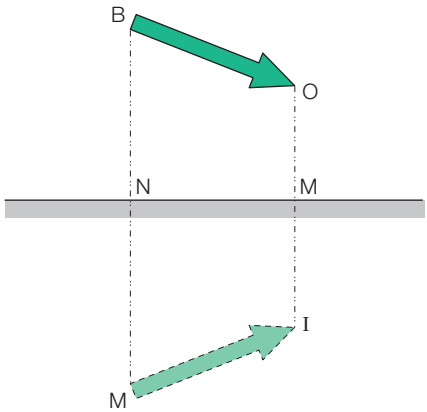


Figure 21.6

Figure 21.6 shows IM as the image of an arrow, OB. The last two activities would have shown that the image point formed is as far behind the mirror as the object point is in front of it. It follows then that if OI meets the mirror at M and BM at N, then

OM = MI, and

BN = NM.

Using these equalities and appropriate geometrical constructions, it can be proved (by congruent triangles) that

OB = IM

This result tells us that the image of an object formed by a plane mirror is of the same size as the object.

image size = object size

MATHEMATICS: congruent triangles

An alternative proof of this result (using the congruency of triangles) is given in the caption of figure 21.7.

The size of the image

It is not easy to measure accurately the size of the image produced by a plane mirror because of its virtual nature. (What makes it so difficult?) Using geometry and the laws of reflection, however, we can show that the image has the same size as the object, either with the help of figure 21.6 or by using the argument in the caption of figure 21.7.

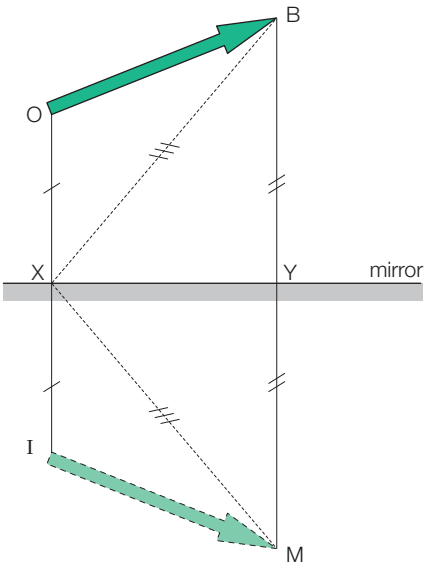


Figure 21.7 The triangles XYB and XYM are congruent (two sides and the included angle). The triangles OBX and IMX are also congruent (again two sides and the included angle). It follows that the corresponding sides OB and IM of triangles OBX and IMX are equal. Thus, the size of the image is equal to the size of the object.

IT Q1

An object is placed between two plane mirrors inclined at 90° to each other (figure 21.8). Show by drawing a ray diagram that three images are produced by the mirrors.

The orientation of the image

MATHEMATICS: plane geometry – congruent triangles

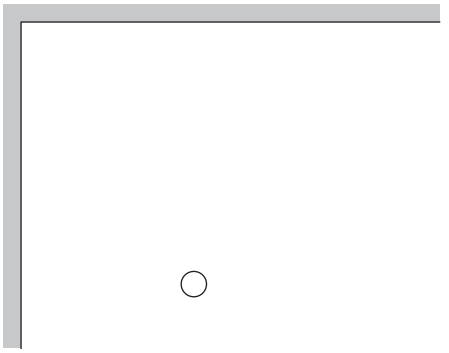


Figure 21.8

Figure 21.9 shows a plane mirror forming an image of an arrow OB. The rays drawn from the ends of the arrow are reflected by the mirror. These diverging reflected rays from O and B enter the eye, which sees the image 'in the mirror'. As the figure shows, the image is not 'in' the mirror at all, but behind the mirror and as far behind the mirror as the object is in front of it. However, you will notice that, after reflection, light from the right-hand side of the object (the incident red rays) appears to come from the left-hand side of the image (the reflected red rays). It seems that what was on the right of the observer (the observer being where the eye is positioned), like the point O on the object now appears in the image on the left-hand side of the observer (at the point I). Another way of saying this is to say that the image is **laterally inverted**.

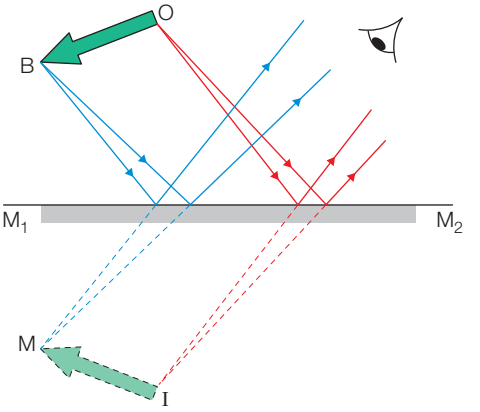


Figure 21.9 The image produced by a plane mirror is laterally inverted.

laterally inverted

A few points to note about lateral inversion

The inversion of the image in figure 21.9 is about the line M_1M_2, called the 'mirror line'. The mirror line is the line in which the plane of incidence (the plane containing the incident ray, the reflected ray and the normal (see page 308) meets the plane of the mirror. A plane mirror always produces lateral inversion about this line. We might regard the image of OB as having been obtained by carrying out a transformation in which OB was turned through 180° about the mirror line M_1M_2.

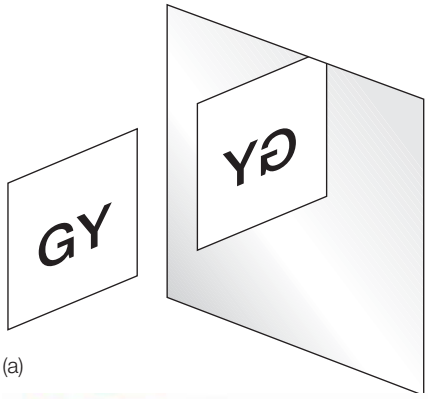
MATHEMATICS: rotation

For example, figure 21.10 (a) shows a card with the letters 'GY' which has been placed at right angles to a mirror. The lettering in the image is laterally inverted about the line at which the plane of the card (which is the plane of incidence) meets the plane of the mirror.

mirror image relationship ›

IT Q2

I am standing in the doorway of a shop and above the door is a sign reading 'CLOSING DOWN'. Directly opposite this sign across the street is a large plane mirror and in this mirror is an image of this sign. How will this sign appear in the mirror?



(a)



(b)

Figure 21.10 There is lateral inversion about the mirror line in each of the cases (a) and (b).

Figure 21.10 (b) shows a boy standing in front of a plane mirror, with his right hand raised. The image of the boy in the mirror has the left arm raised. Again, the image of the boy's right hand has been laterally inverted, (turned through 180° about the mirror line in the plane of the mirror) and appears as his left hand in the mirror. This is what is commonly referred to as 'a **mirror image relationship**' – a rotation of 180° about an imaginary mirror line.

To take an example of this: when you hold your palms open in front of you, they have a mirror image relationship with each other. If you put a plane mirror between the palms, one palm is the image of the other. Put a mirror to the right of your right palm and you get an image which is the left palm. Put a mirror to the left of your left palm and you get the right palm as image. We will find an understanding of this relationship very useful and almost crucial when we come to inversion produced by converging lenses in a later chapter.

Do not confuse lateral inversion with ordinary inversion. Ordinary inversion is always a rotation through 180° about a horizontal axis. You will get an ordinary inversion (where the image is upside down) from a plane mirror only if the mirror line is horizontal, that is, if the mirror is in a horizontal plane. So if someone stands on the edge of a horizontal mirror, the image of that person will be inverted in the normal sense of the word. The image will at the same time show lateral inversion. If the mirror is in a vertical plane, however, it will not produce ordinary inversion or an upside down image, but it will (as it will always do) show lateral immersion. What you will get is, as expected, lateral inversion where 'object left is image right'. You will get an upright image. What was below in the object is still below in the image. This follows from the fact that the image is always opposite the image 'across' the mirror.

The meaning of the term 'upright' (or 'erect') in ray optics

When used in ray optics, the term 'upright' (or 'erect') always means 'oriented in the same way as the object'. So if a device forms an image of a person standing on his or her head, the image will show the person standing on his or her head. This image in ray optics is described as 'upright' (or 'erect'). If the object is on its feet, the image is also on its feet. If, on the other hand, the object is on its head, then so is the image.

Summary of the features of the image formed by a plane mirror

The features of the image formed by a plane mirror are that it is:

1 as far behind the mirror as the object is in front of it, or:
object distance = image distance;
2 virtual;
3 the same size as the object;
4 upright;
5 laterally inverted.

Point 4 means that when we stand facing a plane mirror, the image is the same way up as the object, that is, it is upright (or 'erect').

Point 5 means that what is right on the object appears as left in the image and vice versa.

ITQ3

Figure 21.11 shows the image of a clock face seen in a mirror. What time is shown on the actual clock if the clock was one which had no numbers on its face, as in figure?

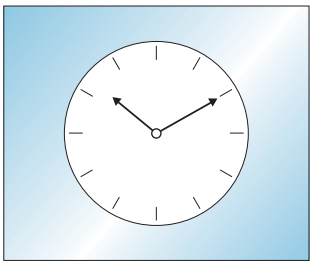


Figure 21.11

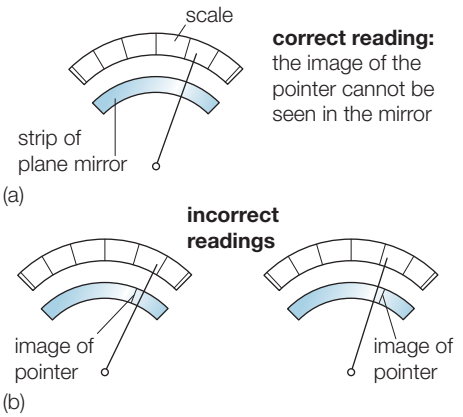


Figure 21.12 (a) Correct reading. (b) Incorrect readings; the eye is to the left or right of the pointer and the reflected image of the pointer can be seen.

Some uses of plane mirrors in physics

Plane mirrors have many uses, for example as rear-view mirrors for bikes or cars, in periscopes and kaleidoscopes, and even in the scales of moving-coil ammeters and voltmeters as described below.

Improving the accuracy of reading a meter scale

Figure 21.12 shows how a strip of plane mirror alongside the scale of an instrument helps the experimenter to avoid a parallax error in reading the position of the pointer. For a correct reading, the experimenter's line of sight is perpendicular to the scale and the image of the pointer cannot be seen. This topic was discussed in chapter 3 in the section 'reducing parallax errors'.

Chapter summary

- Reflection of light is the bouncing back of light at an interface between two different media.
- The laws of reflection are as follows:
  - The incident ray, the reflected ray and the normal at the point of incidence are all in the same plane.
  - The angle of incidence is equal to the angle of reflection.
- The image formed by a plane mirror is always:
  - virtual;
  - upright;
  - of the same size as the object;
  - the same distance away from the mirror as the object;
  - laterally inverted, that is, rotated through 180° about the 'mirror line'.

Answers to ITQs

ITQ1 The object, O, will have three images, A, B and C, in the two mirrors. The red rays show the formation of image A. The blue ray reflected at X gives the formation of the image labelled B (see figure 21.13). The green ray that strikes the other mirror produces the image labelled C. The fainter lines are the extensions of the two mirrors.

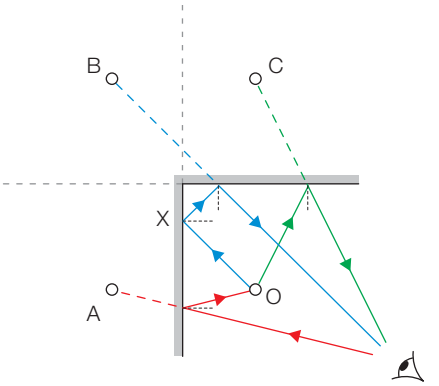


Figure 21.13

ITQ2

CLOSING DOWN	CLOSING DOWN	CLOSING DOWN

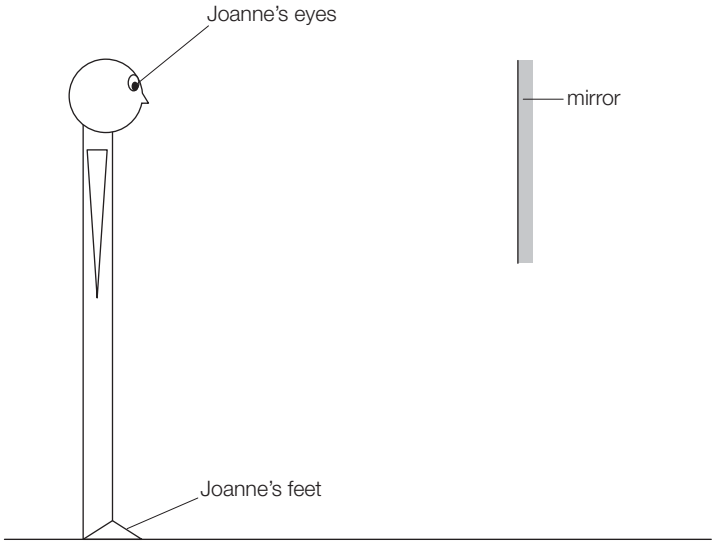
ITQ3 The time is 1:50 or ten-to-two.

Examination-style questions

1 Copy the passage that follows, using the correct words or phrases from those given in brackets to make the passage correct.

The image produced by a plane mirror is always (**virtual** / **real**) and (**upright** / **inverted**). It is (**the same distance away from** / **closer to** / **further away from**) the mirror compared with the object. The image is always (**virtual** / **real**) because the rays leaving the mirror after reflection always (**converge** / **diverge**). Such an image (**can** / **cannot**) be obtained on a screen. It is laterally inverted.

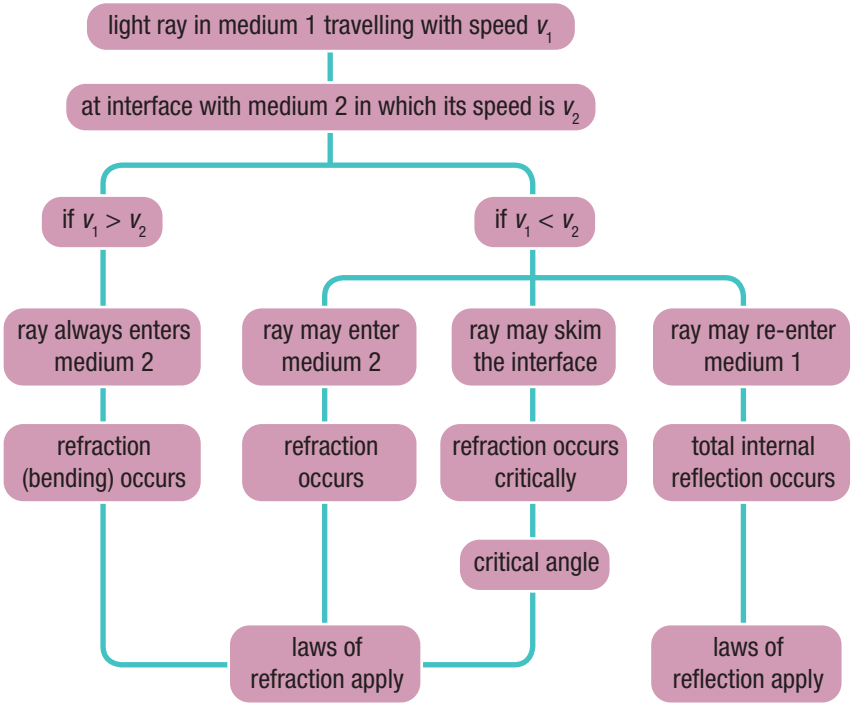
2 Joanne finds that, if she stands in front of a certain mirror that is at a certain height above the floor, she can just see an image of her entire body. She also finds that the height of the mirror is one-half her own height. If Joanne is 164 cm tall and her eyes are 14 cm below the top of her head, show by drawing rays on the diagram that the bottom of the mirror must be 75 cm above the floor.



22 Refraction of Light and its Relation to Colour

By the end of this chapter you should be able to:

- recall the meaning of refraction and the conditions in which refraction takes place
- state the laws of refraction
- perform experiments to test the laws of refraction
- draw diagrams representing the passage of light rays through rectangular blocks and triangular prisms
- recall that the passage of a ray of light:
  - through a parallel-sided, transparent medium may result in the lateral displacement of the ray
  - through a triangular transparent prism may result in deviation of the ray
  - through a triangular transparent prism may result in dispersion of that ray
- give examples of observations that indicate that light can be refracted
- define the refractive index of a medium
- use Snell's law in the solution of numerical problems
- explain with the aid of diagrams what is meant by the critical behaviour of a ray when passing from one medium to another
- explain with the aid of diagrams what is meant by critical angle and total internal reflection
- recall the conditions necessary for a ray to behave critically
- recall the conditions necessary for total internal reflection to occur
- perform calculations involving critical angle and total internal reflection
- explain what is meant by the dispersion of white light to produce a spectrum
- recall what an impure spectrum is and how it may be produced
- recall what a pure spectrum is and how a pure spectrum may be produced
- recall how the spectral colours may be recombined to produce white light
- recall what is meant by an additive primary colour
- recall that the three additive primary colours for mixing light are red, blue and green
- predict what colour will result when light of two different primary colours are mixed
- recall the meaning of the term 'secondary colours'
- recall that the colour of an object is the colour of the light that is reflected by that object
- predict what the colour of an object will be when viewed in light of a particular colour



Introduction

Refraction of light is responsible for many of the natural sights that we enjoy, such as rainbows and mirages. In fact, without the refraction of light by the lenses of our eyes, we would not be able to see at all! Much of our present scientific knowledge has been gained using light-refracting instruments such as the microscope and the telescope.

In this chapter we will describe refraction, the laws that govern it and some of its simple, but significant, consequences.

refraction ›

Refraction may be defined as the bending of light on passing from one medium to another where the speed is different.

angle of incidence ›

The **angle of incidence** is the angle between the incident ray and the normal drawn at the point of incidence, this point being the point where the incident ray strikes the interface between the two media. The angle of refraction is the angle between the refracted ray and the normal. This interface between the two media is called the **refracting surface**. Another term used to represent this surface is the **boundary**.

refracting surface ›
boundary ›

We have seen previously that a ray of light may be represented by the line joining any two points in space since; in fact, light rays are moving in all directions imaginable. Thus, to define a light ray passing through a rectangular block, we may use two small marks on opposite surfaces of the block, as in Practical activity 22.1.

Practical activity 22.1

To show that light may bend on passing out of a glass block

You will need:

- rectangular glass or Perspex block
- measuring cylinder
- large nail or optical pin
- retort stand and clamp
- cork
- plasticine.

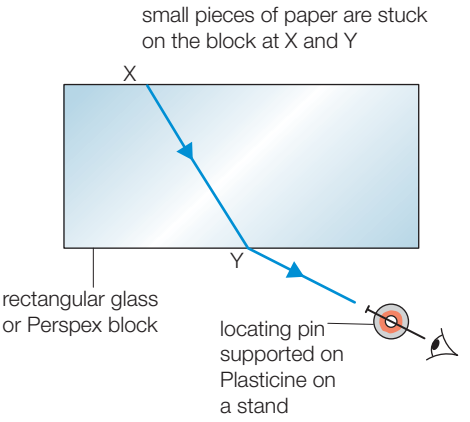


Figure 22.1 A ray of light is bent as it passes from the block to air. The apparatus as seen from above.

Method

1 Stick two small pieces of white paper (about 2 mm × 2 mm) on opposite faces of the block, as shown in figure 22.1, so that they are the same distance from the upper surface of the block.

2 Place the block on top of the measuring cylinder.

3 Stick a small lump of plasticine on to the cork and clamp the cork in the retort stand. Place the pin horizontally on the plasticine.

4 Raise the pin in the stand to the level of the two pieces of paper at X and Y on the block.

5 Looking along the length of the pin, turn the pin and the stand until the pin is in line with X and Y as seen through the block. When this has been achieved, you should be able to see only the piece of paper at Y, behind the pin.

6 Place a metre rule on its edge along the line XY.

Questions

1 Does the edge of the metre rule lie on the same line as the pin?

2 What happens to a ray of light when it passes from glass (or Perspex) to air?

Observation

You will find that the rule does not lie along the line of the pin, but makes an angle with the pin.

Conclusion

As a ray following the line XY leaves the block, it bends as shown by the line of the pin. The ray is bent or refracted on leaving one transparent medium (glass or Perspex) for another (air).

The laws of refraction

There are two laws of refraction:

1 The incident ray, the refracted ray and the normal at the point of incidence are all in the same plane.

2 The ratio of the sine of the angle of incidence (i) and the sine of the angle of refraction (r) for any two given media has a constant value, or

$$\frac{\sin i}{\sin r} = \text{a constant for the two given media.}$$

Snell's law

The second law above is known as Snell's law.

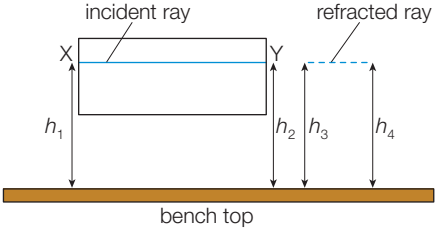


Figure 22.2 Side view of apparatus. Heights h_1, h_2, h_3 and h_4 are measured.

Testing the first law of refraction

Using the apparatus described in Practical activity 22.1, we measure the vertical heights of the points X and Y from the bench top (which is assumed to be horizontal) with the metre rule, and also the vertical heights of the ends of the pin (figure 22.2). If these heights are all the same, within the limits of experimental error, then the line XY, which represents the incident ray, and the pin are in the same plane if the block was horizontal.

Repeat this test for three or four different directions of the 'ray' XY.

Testing the second law of refraction (Snell's law)

The apparatus of Practical activity 22.1 is adapted to look like that of figure 22.3. In addition a drawing board will be required.

1 The sheet of paper is fixed to the drawing board with adhesive tape. The block is placed on a sheet of paper with the two small pieces of paper at X and Y at the lower edge of the block, that is the edge resting on the paper.

2 Having stooped to the level of the block, and with one eye closed, the block is viewed at eye level. An optical pin, P_1, is stuck upright in the sheet of paper so that it is in line with the two small pieces of paper at X and Y when viewed through the block. This pin should be at least 10 cm from the block for good accuracy.

3 Another pin P_2, is stuck upright in the paper in such a position that it, too, is in line with X and Y (as seen through the block) and also with the first pin P_1. This second pin should be at least 10 cm from the first.

4 Draw small circles round the pinpoints, P_1 and P_2, and remove the pins.

5 Mark the positions of X and Y on the sheet of paper.

6 Draw the outline of the block on the paper, making sure that the pencil point used to do this is directly under the edge of the block. Remove the block from the paper.

7 Produce the line P_2P_1 to meet the outline of the block at Y. Draw the line XY.

8 At Y, construct a normal (NYM in figure 22.3) using a protractor. Measure the angles θ_1 and θ_2 that 'rays' XY and P_1P_2 respectively make with the normal NM.

9 Record these values in the format shown in table 22.1.

10 By changing the position of Y on the block, make the angle θ_1 have values between 10° and 40°, at intervals of 5°. The positions for Y could be marked first on the paper using a protractor, the angles being marked with their values. The block is then put in position and the small piece of paper then placed at Y.

11 To avoid congestion of marks on the paper, angles of incidence 10°, 20°, 30° and 40° are investigated with the block in one position on the sheet of paper, and the block is moved to a new position to investigate the remaining angles.

12 When the values of θ_2 have been obtained and recorded for all the values of θ_1 used, plot a graph of sin θ_2 against sin θ_1.

If the points of the graph are on a straight line, then the ratio of sin θ_2 to sin θ_1 is constant, and Snell's law is confirmed.

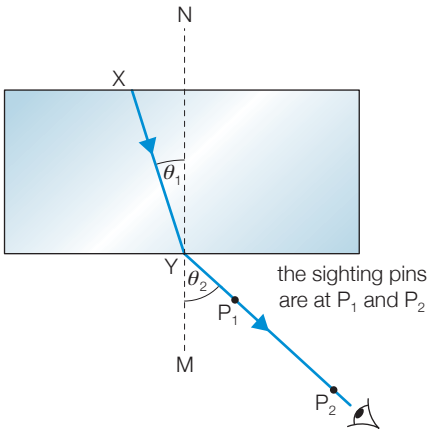


Figure 22.3 Verifying Snell's law.

Table 22.1

θ_1/°	θ_2/°	sin θ_1	sin θ_2

Tracing the path of light rays through rectangular blocks and triangular prisms

In order to do this, we use optical pins to define the path of rays outside the block or prism, as shown in Practical activities 22.2 and 22.3 on the following pages.

Practical activity 22.2

Passage of light rays through a rectangular block

You will need:

- rectangular glass or Perspex block
- four optical pins
- large sheet of plain white paper
- protractor
- pair of dividers
- drawing board
- adhesive tape.

Method

1 Having fixed the paper on the drawing board with adhesive tape, draw a long line, AB, about 10 cm from the long edge of the paper and parallel to it.

2 Using the protractor, draw a normal to the line AB.

3 At the point of intersection of the normal with the line AB, construct an angle to the normal of about 50°.

4 Place two pins at least 6 cm apart on the line that defines the angle just constructed with the normal. Label these two points P_1 and P_2, as shown in figure 22.4. Make sure the pins are vertical.

5 Place the rectangular block on the sheet of paper, with one of its long edges directly over the line AB, as shown in the figure.

6 Stoop to the level of the block, and, with one eye closed, look through the block and stick the other two pins into the paper so that they are in line with P_1 and P_2 as seen through the block.

7 Draw small circles round the pins, and label the points P_1, P_2, P_3 and P_4 as in the figure.

8 Draw a line, CD, along the other refracting surface of the block. Remove the block.

9 Draw a line from P_4 to P_3 and produce this line to meet CD at U.

10 Draw a line from P_1 to P_2 and produce this line to meet AB at V.

11 Join U to V.

12 Produce P_1P_2 across the outline of the block as shown by the dashed line in figure 22.4.

13 From the points P_3 and P_4, draw normals to P_1P_2 to meet the extension of line P_1P_2 at Y and Z.

14 Measure P_3Y and P_4Z.

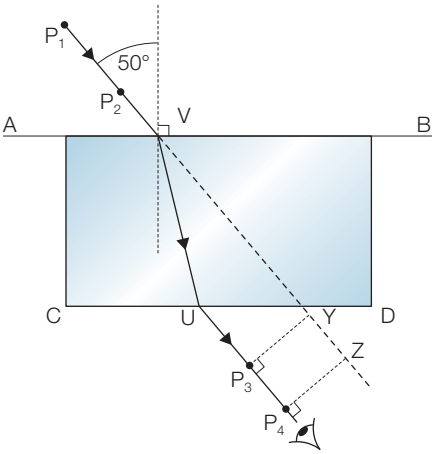


Figure 22.4 Tracing the path of rays through a block.

Observations

$P_1P_2VUP_3P_4$ represents the path of a ray passing through the block. Of course, if the four pins are viewed from the other side of the block, they will again appear to be in line (the reversibility principle). (Check that this is so.)

The path of the ray before it entered the block (the incident ray) is shown by P_1P_2. The path of the ray on leaving the block (the emergent ray) is shown by P_3P_4, and the distances P_3Y and P_4Z should be equal if the sighting of the rays and the positioning of the sighting pins were carefully done.

Conclusion

If the lengths P_3Y and P_4Z are equal, then the initial and final directions of the ray are parallel. This means that the ray is displaced to one side (laterally) as a result of passing through the parallel-sided glass (or Perspex) block

We can sum up the observations in Practical activity 22.2 by noting that, when a ray passes through a parallel-sided glass block, the emergent ray has the same direction as the incident ray but it has been shifted sideways. This sideways shift is called **lateral displacement**.

lateral displacement ›

Latin *latus* means 'side'.

It is easy to show practically that the amount of displacement produced by the block depends in the angle of incidence of the ray and, also, on the width of the block. It can be shown that the lateral displacement produced varies directly as the width of the block if the angle of incidence is held constant. We will use this fact a little later when we come to study lenses and how they work to produce images.

Of course, this can easily be proved by using some elementary geometry together with Snell's law.

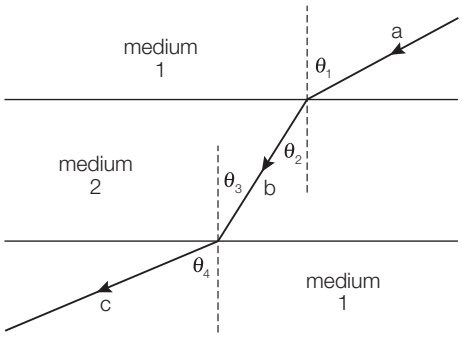


Figure 22.5

Look at the diagram of figure 22.5. The rays marked a, b and c are the incident ray, the transmitted ray and the emergent ray. θ_1 and θ_2 are related by Snell's law in that $\sin \theta_1 / \sin \theta_2$ = a constant, k.

Since the refracting surfaces are parallel, then $\theta_2 = \theta_3$, and for the refraction taking place at the second surface, we can write that

$$\frac{\sin \theta_4}{\sin \theta_3} = \text{the same constant, } k$$

It follows that $\dfrac{\sin \theta_1}{\sin \theta_2} = \dfrac{\sin \theta_4}{\sin \theta_3}$

But $\theta_2 = \theta_3$

And so it follows that $\theta_1 = \theta_4$ and that, since the normals are parallel to each other, then the emergent ray c is parallel to incident ray a.

MATHEMATICS: geometry

The reversibility principle

If the emergent ray, c, were reversed to become the incident ray, by Snell's law ray b would remain the transmitted ray travelling on the opposite direction. Since, by Snell's Law, the ratio ($\sin \theta_4 / \sin \theta_3$) is also equal to k for the refraction at the second surface, and since $\theta_3 = \theta_2$, then it follows that $\theta_4 = \theta_1$ and ray a will be parallel to ray c. This means that the new transmitted ray would follow the same path as the old. Clearly, also, by a similar reasoning the new emergent ray would follow the path of the previously incident ray, a.

This is an illustration of what is called 'the reversibility principle' relating to light rays.

The reversibility principle states that if, in passing through an optical system, a light ray follows a certain path, it would follow the very same path, but in the opposite direction, if it were reversed through the optical system.

Whereas a ray passing through a rectangular block is laterally displaced, light passing through a triangular prism is deviated by the prism. This is demonstrated in the next practical activity, Practical activity 22.3. The deviation produced by a triangular prism is also useful in understanding the action of lenses.

Note also that it does not matter what is the medium on either side of the block. What is important is that the medium should be the same on both sides and the refracting surfaces be parallel.

Practical activity 22.3

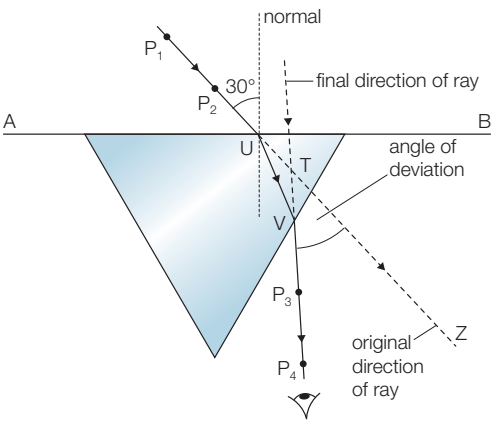


Figure 22.6 Tracing the path of rays through a triangular prism.

Do not confuse lateral displacement and deviation.

Deviation of light rays by a triangular prism

You will need

- equiangular triangular prism
- four optical pins
- large sheet of plain paper
- protractor
- drawing board
- adhesive tape.

Method

1 Fix the sheet of paper on the drawing board with the adhesive tape. Draw a long line, AB, about 10 cm from the long edge of the paper and parallel to it. Mark a point on this line, as shown in figure 22.6.

2 At this point, using the protractor, draw a normal to the line AB.

3 At the point of intersection of the normal with the line AB, construct an angle to the normal of 30°.

4 Place two pins at least 6 cm apart on the line P_1P_2, as shown in figure 22.6, which makes an angle of 30° with the normal. Make sure the pins are vertical.

5 Place the triangular prism on the sheet of paper, with one of its refracting surfaces directly over the line AB, as shown.

6 Stoop to the level of the prism and look through it with one eye. Stick the other two pins into the paper, at least 10 cm apart, at P_3 and P_4, so that they are in line with P_1 and P_2 when seen through the block.

7 Mark the positions of the pins with small circles, and label the points P_1, P_2, P_3 and P_4 as in the figure.

8 Draw in the outline of the prism, making sure the lines are drawn directly under the corresponding faces. Remove the prism.

9 Draw a line from P_1 to P_2, to show the path of the incident ray, and produce this line to some point Z as shown in figure 22.6. Mark the point U where the incident ray meets the first refracting surface of the prism.

10 Draw a line from P_4 to P_3, to show the path of the emergent ray, and produce this line back. Denote the point where it meets the prism as V, and the point where it meets line P_1P_2Z, the direction of the incident ray, as T, as shown in figure.

11 Measure the angle P_3TZ.

12 Draw a line from U to V.

Observations

The path of the ray through the prism is shown by the line UV. Since the path of the ray outside the prism is different from the original path, the ray has been deviated, that is, it has been made to follow a different course. The angle through the ray has been taken off its original course is the angle of deviation of the incident ray. This angle is the angle P_3TZ.

Conclusion

As a result of passing through the prism the ray P_1P_2 suffered a deviation of the size of angle P_3TZ. This angle is called the angle of deviation of the ray P_1P_2.

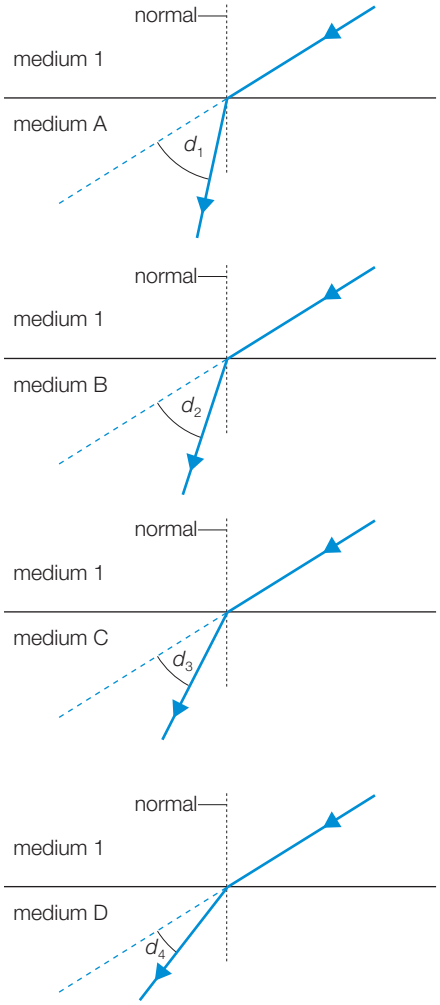


Figure 22.7 Comparing refraction in four different media A, B, C and D. The angle of incidence is the same in each case. The angles marked d show how far each ray is refracted from its original path.

refractive index ›

Lateral displacement and deviation are very different effects. Whereas in the first case the ray maintains its direction but its position changes, in the latter case the ray changes its direction. If there was a similarity between the two, it would be that in both cases the size of the effect depends on factors related to the object that produces it. The amount of lateral displacement produced depends on the width of the rectangular block. In similar manner the amount of deviation the ray suffers depends on the angle between the refracting faces of the prism (called the refracting angle of the prism). The greater the width of the block, the greater the lateral displacement. In like manner, the greater the angle of the prism, the greater is the deviation produced.

The refractive index: comparing refraction in transparent media

Light bends on entering a different medium because its speeds in the first and second media are different. We would expect, therefore, that when rays of light leave one medium to enter other media where the speed has different values, there will be different degrees of refraction (or bending) corresponding to the different speeds in the second media.

If we investigate this, we must be careful, however, to make sure that all the factors which affect the amount of refraction are kept the same, except the one factor whose effect we are examining (controlling variables?). So the first medium must be the same and the angle of incidence in that medium must also be the same. We would then measure the amount of bending that occurs in different second media. Figure 22.7 shows the possible results of such an experiment.

The amount of bending (or deviation) that occurs is clearly different from medium to medium. We find that, in this case where the incident ray is in medium 1, the most refractive medium seems to be A, where the deviation is greatest, followed by B, C and D in that order.

We could not say more than that, however. If d_1 was twice as large as d_2, we could not state that medium A was twice as refractive as medium B, because if we used a different reference medium, the ratio d_1/d_2 might not be the same.

In fact, we measure refracting ability of a medium using a quantity called the **refractive index**, whose symbol is n. The refractive index of a medium is defined as the ratio of the speed of light in a vacuum to the speed of light in that medium:

$$\text{refractive index of a medium} = \frac{\text{speed of light in a vacuum}}{\text{speed of light in the medium}}$$

or

$$n_{\text{medium}} = \frac{c_{\text{vacuum}}}{c_{\text{medium}}}$$

where c is the velocity of light in a vacuum.

Note that we are using the letter 'c' to represent the speed of light. This letter is the only one that is used to represent the speed of light.

Since a vacuum would have no matter to impede the light as it passes through, the light would have its greatest speed in a vacuum. However, since:

(i) air is all around us and, at normal atmospheric pressure, the molecules are not close enough together to significantly affect the speed of light passing through it, and

(ii) we cannot readily produce a vacuum in order to find the refractive index of a substance, we use air as our reference medium. The definition we use, in practice, is

$$\text{refractive index of a medium} = \frac{\text{speed of light in air}}{\text{speed of light in that medium}}$$

or, for short, R.I. (refractive index) of the medium, n, is given by

$$n_{\text{medium}} = \frac{c_{\text{air}}}{c_{\text{medium}}}$$

Therefore, if the medium is air, we can see that the R.I. of air = $c_{\text{air}}/c_{\text{air}} = 1.0$. In all our calculations involving refractive index we shall take the R.I. of air to be 1.0.

Remembering that light is a wave motion and that the treatment applied to sound, a wave motion, a few chapters ago could also be applied to light, we can presume that, if a ray of light made an angle of incidence θ_{air} in air (figure 22.8) and, on entering a second medium, the angle of refraction there was θ_{med}, then, from Snell's law, we could say that

$$\frac{c_{\text{air}}}{\sin\theta_{\text{air}}} = \frac{c_{\text{med}}}{\sin\theta_{\text{med}}}$$

and so

$$\frac{c_{\text{air}}}{c_{\text{med}}} = \frac{\sin\theta_{\text{air}}}{\sin\theta_{\text{med}}}$$

But $\frac{c_{\text{air}}}{c_{\text{med}}}$ = R.I. of the medium = n_{med} (by definition)

and so it follows that $n_{\text{med}} = \frac{\sin\theta_{\text{air}}}{\sin\theta_{\text{med}}}$

The refractive index of a medium is therefore given by

$$n_{\text{med}} = \frac{\sin(\text{angle in air})}{\sin(\text{angle it the medium})}$$

the angle in question always being the angle which the rays make with the normal.

This formula giving an expression for refractive is not the definition of refractive index, but a convenient equivalent which is easy to use, since angles can easily be measured. We cannot as easily measure the speed of light in any medium.

A brief note about the term 'optical density'

Sometimes the term 'optical density' is used to represent refractive index and you might read of 'the optical density of medium A' being greater than 'the optical density of medium B'. We strongly recommend that you do not use this term, since it has nothing to do with density as used in physics, and refraction has nothing to do with density of any sort. In this book the preferred term to represent a medium of high R.I. is 'slow medium' (and not high optical density) and to represent a low R.I. of a medium it is better to use 'fast medium' and not low optical density. The preferred terms represent the truth about the speeds of light in media except that the terms 'fast' and 'slow' are attached to the media rather than to the light itself.

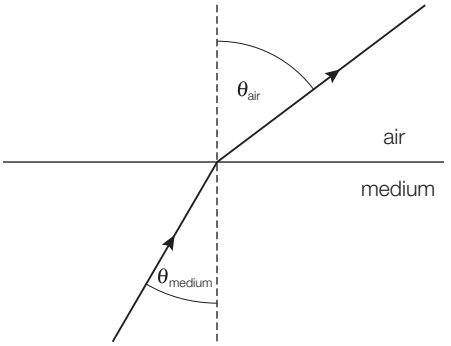


Figure 22.8

Reversibility principle applied

It is very important to point out that this formula will hold whether the ray is passing from air into the medium or from the medium into the air. This must follow from Snell's law. As long as the media remain the same A and B, say, then the ratio of the sines of the angles concerned will remain the same whether the ray is proceeding from medium A to medium B or vice versa. This is based on 'the reversibility principle' discussed earlier which, simply expressed, states that:

If a light ray follows one path in getting from point A to point B in an optical system, it would follow exactly the same path if it were reversed and started from B in an effort to get to A.

Note, too, that it is the ratio of the *sines* of the angles that is constant and *not* the ratio of the angles themselves.

Worked example 22.1

The speed of light in a certain medium is $2.2 \times 10^8\,\text{m s}^{-1}$. Calculate the refractive index of the medium if the speed of light in air is $3.0 \times 10^8\,\text{m s}^{-1}$.

Solution

The R.I. of the medium $= \frac{c_{\text{air}}}{c_{\text{med}}}$

$$= \frac{3.0 \times 10^8\,\text{m s}^{-1}}{2.2 \times 10^8\,\text{m s}^{-1}}$$

= 1.4 (no units) and to 2 sig. figs (why 2 significant figures?)

Worked example 22.2

A light ray leaves the same medium as in worked example 1 and enters air. The angle of incidence in the medium is 40°. Calculate the angle of refraction in the air.

Solution

From the definition of the R.I. of the medium,

Note that the angle in air must be in the numerator

$$\text{R.I. of the medium} = \frac{\sin(\text{angle in air})}{\sin(\text{angle in the medium})}$$

$$\text{Or } \frac{30}{22} = \frac{\sin(\text{angle in air})}{\sin 40°}$$

($\frac{30}{22}$ being the actual value of the R.I. of the medium.)

This gives

$$\sin(\text{angle in air}) = \sin 40° \times \frac{30}{22}$$

$$= 0.876.....$$

$$= \sin 61.2°$$

So the angle of refraction in the air or the (angle in air) = 61°.

Note very carefully that the light has left a medium of higher R.I. to enter another of lower value and it bends away from the normal. We will shortly show that this will always happen – that in leaving one transparent medium for another, the ray will make the larger angle with the normal in the medium with the smaller R.I.

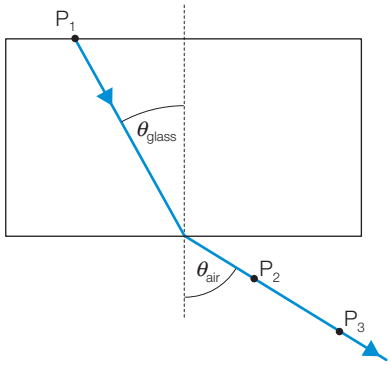


Figure 22.9 Finding the refractive index at a glass–air interface.

Finding the refractive index of a glass block by experiment

For a solid in the shape of a rectangular block, the method is to trace the path of rays through the block using optical pins. The method is the same as that described in Practical activity 22.1. A number of pairs of values of the two angles θ_{air} and θ_{glass} (as shown in figure 22.9) can be obtained. A graph of sin θ_{air} against sin θ_{glass} is plotted.

We use the equation

$$n_{glass} = \frac{\sin \theta_{air}}{\sin \theta_{glass}}$$

Therefore, sin $\theta_{air} = n_{glass}$ sin θ_{glass} and the graph of sin θ_{air} against sin θ_{glass} will have a slope of value n_{glass}, since n_{glass} is a constant.

To find the refractive index of a liquid, the same method can be used if a parallel – sided transparent tank can be found to store the liquid. It would be an advantage if the walls of the tank were as thin as possible, since the lateral displacement of the light rays on entering and leaving the tank would introduce errors in the angles measured – the thicker the walls, the greater the error.

The general formula for refraction at an interface between two media

Consider a light ray leaving a medium 1 at an angle of incidence θ_1 and entering a different medium, medium 2, where the angle of refraction is θ_2.

By Snell's law, $\frac{\sin \theta_1}{c_1} = \frac{\sin \theta_2}{c_2}$ (equation 1)

Multiplying both sides of the equation by c_{air}, we have

$$\sin\theta_1 \times \frac{c_{air}}{c_1} = \sin\theta_2 \times \frac{c_{air}}{c_2} \qquad \text{(equation 2)}$$

Remember that (c_{air}/c_1) = R.I. of medium 1 = n_1

and (c_{air}/c_2) = R.I. of medium 2 = n_2, we have from equation 2

$$\sin\theta_1 \times n_1 = \sin\theta_2 \times n_2$$

or therefore

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \qquad \text{(equation 3)}$$

This equation is extremely important. It can be used to solve any problem involving refraction at a boundary between two transparent media. It tells us that for any two given media with refractive indices n_1 and n_2, the product (R.I. × sine of the angle) for medium 1 = (R.I. × sine of the angle) for medium 2.

Remember always that the angle in question is that between the ray and the normal.

This relationship could be extended to any number of media in contact. We take Worked example 22.3 as an example of the application of this most important 'rule'.

Worked example 22.3

A ray of light travels from glass of refractive index 1.5 into water of refractive index 1.33. The angle of incidence of the ray in glass is 30.0°. Calculate the angle of refraction in the water.

Solution

Using equation (3) above, we have

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

substituting, we have (taking medium 1 to be the medium in which the ray is first present)

$$1.5 \times \sin 30.0° = 1.33 \times \sin\theta_2$$

$$\text{or} \quad \sin\theta_2 = 1.5 \times \tfrac{1}{2} \times \frac{1}{1.33}$$

giving $\sin\theta_2 = 0.5639$

and $\theta_2 = 34.3°$

The angle of refraction of the ray in the water is 34.3°.

ITQ1

A ray of light leaves a small object, M, at the bottom of a vessel containing a layer or oil resting on a layer of water (figure 22.10). The diagram shows the path of the ray of light as it travels through the liquids. The refractive indices of water, oil and air are respectively, n_w, n_o and n_a, and the speeds of light in the three media are respectively c_w, c_o and c_a. It is known that $n_w > n_o$ and, of course that both indices are greater than that of air. Write down:

(i) the speeds of the ray in the three media in ascending order of size;

(ii) the refractive indices of the three media in ascending order of size.

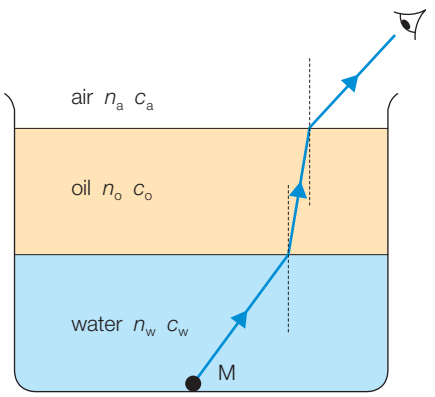


Figure 22.10

Equation (3) states that

$$\text{Sin}\,\theta_2 = \sin\theta_1 \times (\tfrac{n_1}{n_2})$$

The size of the angle θ_2 therefore depends on the ratio (n_1/n_2).

We interpret this to mean that if the ratio $(n_1/n_2) > 1$, then $\sin\theta_2 > \sin\theta_1$ and $\theta_2 > \theta_1$.

Thus, to take an example, if the ray leaves glass (R.I. = 1.50) and enters water (R.I. = 1.33), the important ratio (n_1/n_2) = 1.50/1.33 and this ratio is greater than 1.

So $\sin\theta_2$ will be greater than $\sin\theta_1$ making θ_2 greater than θ_1, which in turn means that the ray will go away from the normal in the medium of lower refractive index.

We now consider some very common cases that are due to refraction.

1 Why does the bottom of a swimming pool appear shallower than it really is when viewed from directly above? The diagram of figure 22.11 explains the answer.

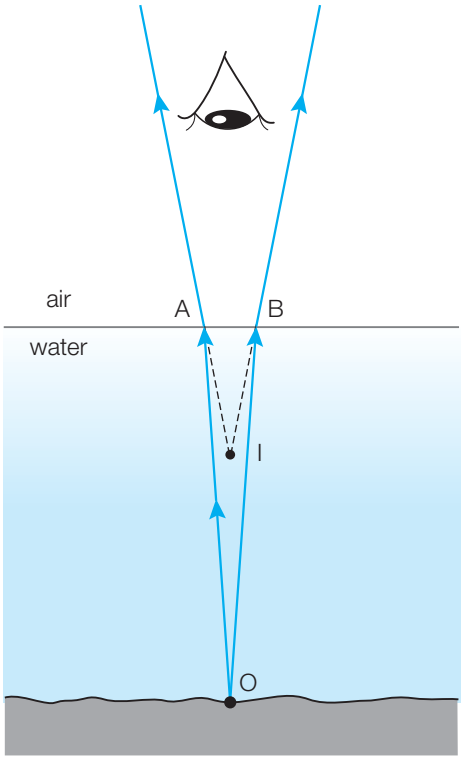


Figure 22.11 The eye sees the image of O at I.

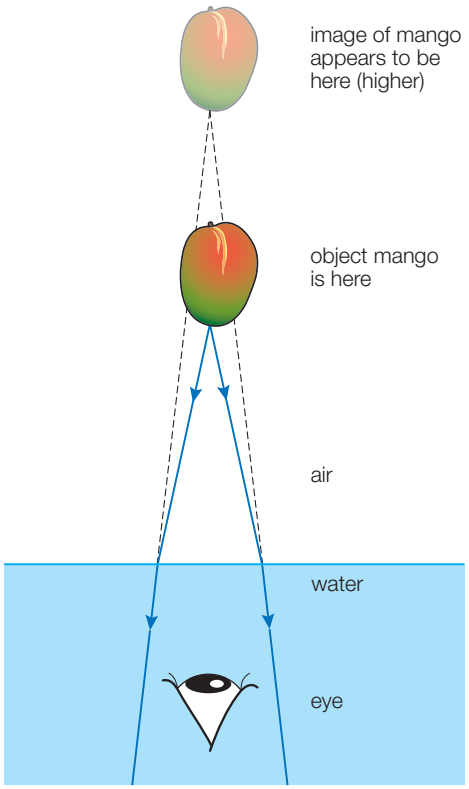


Figure 22.13

2 Why does a point at the far end of the bottom of a swimming pool viewed from the opposite end seem both closer and much less deep than it really is? Study the diagram of figure 22.12 for the answer.

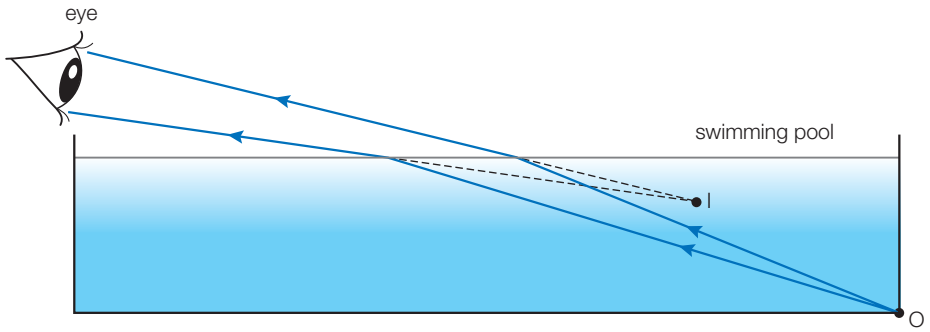


Figure 22.12 The image of the bottom of the pool at O is at I. The eye the bottom of the pool at O seems higher and nearer.

3 If you are in a body of deep water and there are mangoes hanging above the water they would seem further away than they really are. Why is this so? Answer this question with the help of figure 22.13.

Critical angle for two transparent media in contact

What does critical angle mean? How do you calculate its value for those two media?

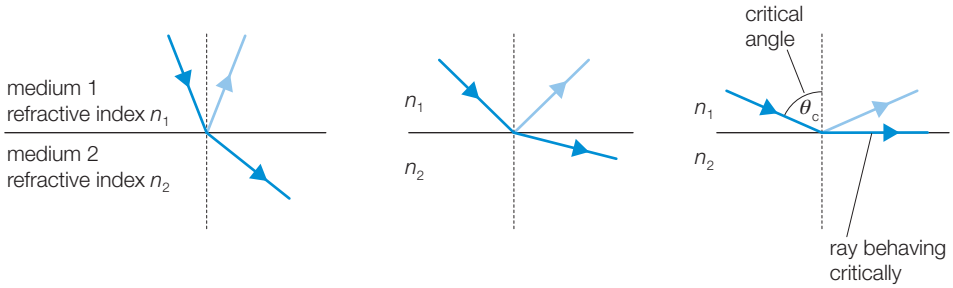


Figure 22.14 As the angle of incidence increases, so does the angle of refraction.

Figure 22.14 shows what happens when a ray of light passes from a medium of refractive index n_1 into another of refractive index n_2, where n_1 is larger than n_2. We expect the angle of refraction in the faster medium to get larger and larger as the angle of incidence in the slower medium increases since, from Snell's law, the ratio of the sines of these angles must remain constant. In other words, if $n_1 \sin \theta_1 = n_2 \sin \theta_2$ at all times

then $\sin \theta_2 = \sin \theta_1 \times \frac{n_1}{n_2}$

and if $(\frac{n_1}{n_2}) > 1$,

then $\sin \theta_2 > \sin \theta_1$

and $\theta_2 > \theta_1$

At a certain point, the angle of refraction in medium 2, θ_2, will become 90°, and the ray will not enter the second medium at all, but will travel along the interface between the two media, at least in theory. It is difficult to observe this in practice. This ray which in theory 'skims' the interface between the two media is called the 'critical ray' and the angle in the first medium which it makes with the normal is called the '**critical angle**'. So, representing the critical angle by θ_c, we can write, as for all cases of refraction,

critical angle

ITQ2

Prove that if a light ray strikes the interface between two transparent media at an angle of incidence of 0°, it will enter the second medium at an angle of refraction of 0°. In other words, a ray that strikes an interface perpendicularly will go straight through without bending. Prove this.

$$n_1 \sin \theta_c = n_2 \sin 90°$$

Since $\sin 90° = 1$, then

$$\sin \theta_c = \frac{n_2 \text{ (smaller R.I.)}}{n_1 \text{ (larger R.I.)}}$$

from which the value of θ_c can be found. It is worth noting that:

1 For there to be criticality (that is, the case where the angle of refraction in the faster medium is 90°), the ray must be leaving a slower medium for a faster one, i.e. the first medium must be of a higher refractive index than the second.

2 The critical angle in a given medium will vary with the nature of the second medium, since its value depends on n_2.

Thus a ray travelling from glass (refractive index 1.5) to water (refractive index 1.33) can behave critically, the critical angle being

$$\sin^{-1} \frac{\text{smaller refractive index}}{\text{larger refractive index}}$$

or

$\sin^{-1}$ means 'the angle whose sine is …'.

$$\sin^{-1} \frac{1.33}{1.5} = \sin^{-1} 0.887 = 62.5°$$

The critical angle for a glass–water interface is 62.5°.

The same ray travelling from glass to air will have a critical angle in air of

$$\sin^{-1} \frac{\text{smaller refractive index}}{\text{larger refractive index}}$$

or $$\theta_c = \sin^{-1} \frac{1.0}{1.5} = 41.8°$$

The critical angle for a ray travelling from glass to air is 41.8°.

The closer the values of the refractive indices of the two media, the nearer the ratio gets to 1 and the nearer $\sin \theta_c$ gets to the value 1. Consequently, as the values of the two indices get closer, the critical angle for that interface gets larger, approaching 90°.

Therefore, for a ray leaving a given medium, the smaller the R.I. of the second medium, the sooner the point of criticality is reached.

Total internal reflection

Look again at figure 22.14. You can see that a small amount of the incident energy is reflected back into the first medium. Most of the incident energy is, however, transmitted as refracted light. Whenever there is refraction there is also some reflection, which we might call 'partial reflection', as only a part of the incident energy is reflected back into medium 1. This is in keeping with what we saw in chapter 21: reflection will take place because the ray is striking an interface between two media with different densities.

If the angle of incidence in the slower medium exceeds the critical angle, all of the incident energy will be reflected back into the slower medium. This is called **total internal reflection** – 'total' because all the incident energy is reflected back into the slow medium and 'internal' because the light remains inside the first (slow) medium. When this reflection takes place, as with reflection of any cause whatever, the laws of reflection are obeyed, as shown in figure 22.15 (b).

total internal reflection ›

Total internal reflection at an interface is only possible if a ray:

1 is incident within a slow medium at an interface with a faster medium, and

2 strikes the interface at an angle of incidence greater than the critical angle for the two media.

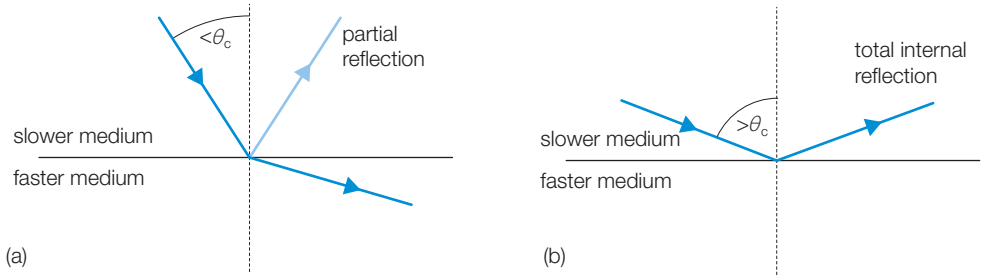


Figure 22.15 (a) The angle of incidence is less than the critical angle. Most of the light energy is refracted; a little is reflected. (b) The angle of incidence is greater than the critical angle and total internal reflection occurs. None of the incident energy is transmitted, all is reflected back into the first medium – total internal reflection.

Total internal reflection is sometimes shortened to T.I.R.

Total internal reflection has found important application in periscopes, binoculars and in light pipes or endoscopes.

We discuss first the periscope.

IT Q3

Assuming that diamond has a refractive index of 2.4, calculate its critical angle at an interface with: (i) air; and (ii) water with a refractive index of 1.33.

The 90°–45°–45° prism

A prism whose angles are 90°, 45° and 45° is called a right-angled isosceles prism. Such prisms are used in many instruments either to turn rays through 90°, as in the periscope, or to turn rays through 180°, as in binoculars.

Use in periscopes: turning a ray through 90°

Figure 22.16 (a) shows a section of a right-angled isosceles prism. The ray of light marked 1 enters the prism perpendicularly (normally) to the face AB, so it passes into the prism without changing course and strikes the hypotenuse face, AC, at an angle of incidence of 45°. This angle is greater than the critical angle of 41.8° for a glass-air interface. The incident ray is therefore totally internally reflected at the face AC and leaves the prism at right angles to the face BC. The incident ray has been deviated through 90° and leaves the prism as ray 2.

You can check this value for the critical angle by using the formula:

$$\text{sine of the critical angle} = \frac{\text{smaller refractive index}}{\text{larger refractive index}}$$

$$\text{sine of the critical angle} = \frac{1.0 \text{ (for air)}}{1.5 \text{ (for glass)}}$$

Ray 3 travelling parallel to ray 1 will behave similarly. All rays that strike a non-hypotenuse face at right angles will be deviated through 90°.

As you can see in figure 22.16 (a), the reflected rays form a laterally inverted image about face AC of the prism, similar to that produced by a plane mirror. The consequences of this reflection are no different from what we met before.

If two prisms are used, one above the other so that the two hypotenuse faces are parallel and facing away from each other, as shown in figure 22.16 (b), parallel rays falling on one of the non-hypotenuse faces of one prism will emerge from the second prism and at a different level. Thus two right-angled isosceles prisms can be used in a periscope, as can two plane mirrors. What the periscope has done, in effect, is to alter the level at which the object is being observed. So when the order 'Up periscope' is given in the submarine what the operator does is to use the periscope to bring an image of the object on the surface of the sea down to the level of the submarine.

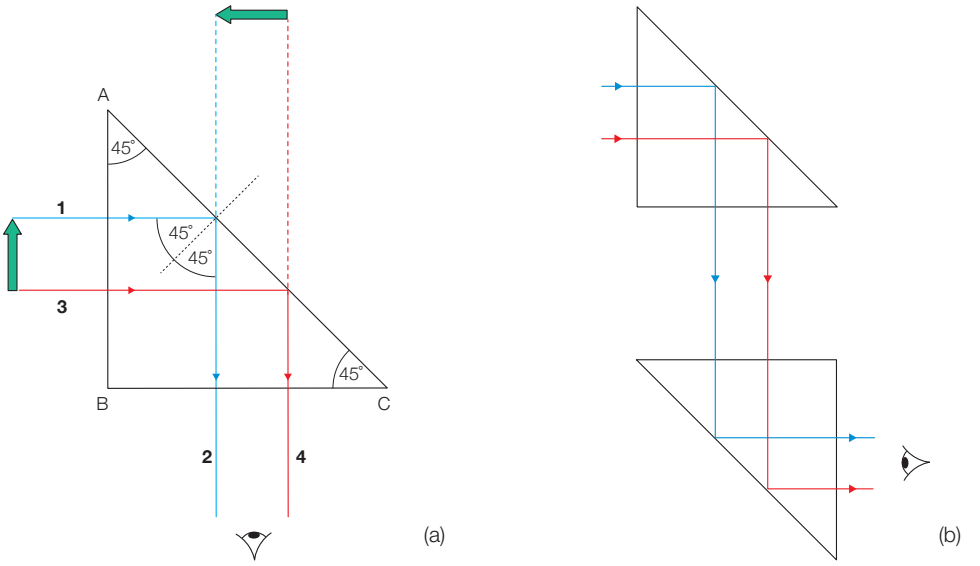


Figure 22.16 (a) The right-angled isosceles glass prism. A light ray that strikes a non-hypotenuse face at right angles is deviated through 90°. The image is laterally inverted. (b) Two such prisms are used in a periscope.

Note that the deviation of the ray here is 180°. Refer to figure 22.17.

Of course, since the hypotenuse face of the prism behaves like a mirror, all the features of the image we discussed for the plane mirror in chapter 21 will apply in this case.

The prism binoculars also use the 90°–45°–45° prism but, rather than using the hypotenuse face to bring about the total internal reflection, it uses the isosceles faces instead. Look at the ray diagram of figure 22.17 (a). Study the diagram and see how the prism inverts the object about an axis passing through the edge formed by the isosceles faces. Study also figure 22.7 (b).

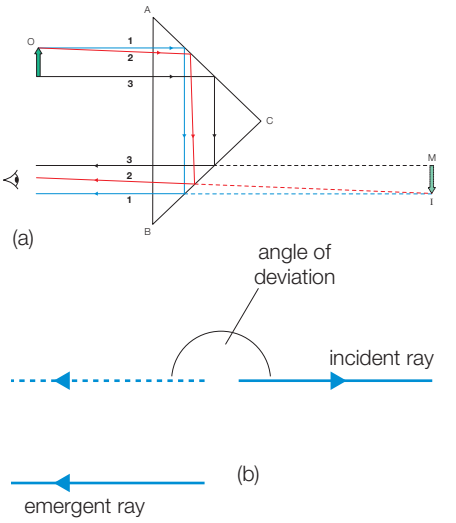


Figure 22.17 Deviation through 180° by a 90–45–45 degrees prism.

Total internal reflection in optical fibres

We saw earlier in the chapter that two transparent media in contact whose refractive indices were very close in value would have a large critical angle. A light ray striking the interface between these media at an angle of incidence larger than this value will undergo total internal refection. This is the principle used in optical fibres.

optical fibre ›

Optical fibres are commonly used for medical diagnosis and also, very importantly, in telecommunications to carry information . An **optical fibre** consists of a core of very fine 'wire' of a material (normally glass or plastic of a special kind) with a refractive index of a certain value and over this core is a layer of a similar material but of a slightly smaller refractive index, called the cladding.

A narrow beam of light enters the core of the fibre at such an angle that it strikes the interface between the two types of glass at an angle of incidence greater than the critical angle for that interface. The light is totally internally

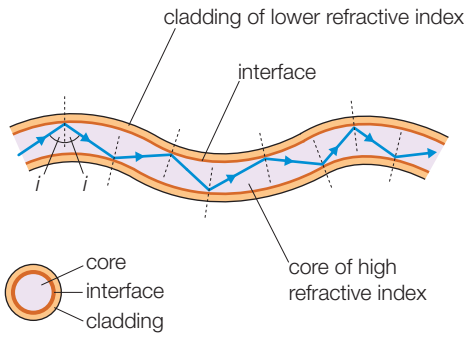


Figure 22.18 Optical fibre.



Figure 22.19 Optical fibres are used in endoscopes to assist in medical diagnosis and treatment.

reflected, strikes the interface again at an angle of incidence too great for refraction to occur and is again totally internally reflected. This process is repeated each time the ray strikes the interface and, in this way, the ray can travel along a considerable length of fibre (figure 22.18).

There are a number of applications of optical fibres: in medicine (in endoscopes, as shown in figure 22.19), in industry and, perhaps most particularly in telecommunication systems, where such fibres are used to transmit electrical signals over large distances with very little loss of signal strength.

An added advantage in telecommunications is that such a fibre can transmit thousands of separate telephone conversations at the same time, since many different beams of light can move along the same medium without interfering at all with one another. Yet another advantage is that the glass and plastics used for making the fibres are much cheaper to produce than the copper cables that were used previously, and very much lighter and more flexible. The diameter of the fibres varies from around 0.1 mm to about 0.5 mm. The light pipes used in medicine and in industry are generally somewhat thicker. The finer the fibre, the more easily can light travel along it by total internal reflection for very long distances.

Dispersion of white light

As we saw in chapter 19, one of the many forms of electromagnetic radiation that reach the earth from the Sun is light. This light we call 'white light'. It is a mixture of the colours we take the spectrum to consist of. We saw that light waves themselves had a range of wavelengths, which we called the light spectrum. It was Isaac Newton who first split light into its component colours, by passing white light through a prism and displaying the spectrum on a screen. This effect is called dispersion. How does a prism disperse light in this way? **Dispersion**, then, is the separation of the colours in **white light** one from another.

dispersion white light ›

CHAPTER 19

We saw in chapter 19 that:

- all forms of electromagnetic radiation (including light) have the same speed in air (strictly, in a vacuum), but only in air and not in other media;
- the frequency of light waves varies within the visible spectrum, with red having the lowest frequency (longest wavelength) and violet the highest frequency (shortest wavelength).

Group discussion

If a narrow beam of white light falls on a glass–air interface in a direction normal to the interface, it will be transmitted without suffering lateral displacement, deviation or dispersion. Can you work out why there will be none of the first two?

If a narrow beam of white light falls on a glass–air interface in a direction normal to the interface,it will be transmitted without suffering lateral displacement, deviation or dispersion. Can you work out why there will be none of the first two?

The reason there will be no separation of the colours is exactly why there is no lateral displacement or deviation. All the colours travel along the same path (but at different speeds) and emerge together along the same path. There will be no difference to the eye and the colours will all be mixed together and appear white. See figure 22.22 (b).

If the ray was incident on the block at an angle, however, the colours would each emerge along a slightly different path, since they would each have a different speed in the block and, by Snell's law (once again!) they would have to have separate angles of refraction. The fastest colour (red) will have the largest angle of refraction (and so the smallest change of course) and the slowest, violet, the smallest angle of refraction (and the largest change of path). This is shown in figure 22.20. To prove this, look at the image of a pin through a glass block from a direction not perpendicular to the refracting surface of the block. The pin appears coloured. This does not happen with a pin that is

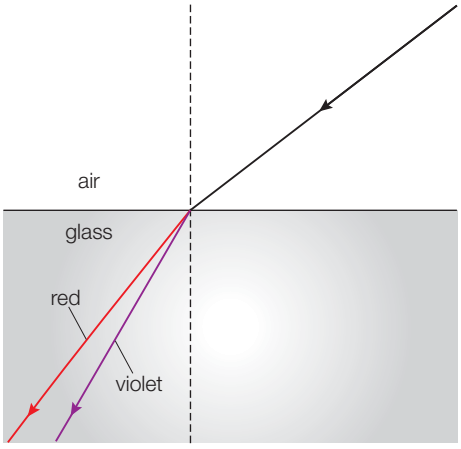


Figure 22.20 Dispersion of white light. Red is the fastest colour and therefore is least refracted. Violet light is slowest and is refracted the most.

viewed in a direction normal to the face of the block.

If these coloured rays are now allowed to fall on another glass-air interface that is not parallel to the first (look at figure 22.21), more separation of colours takes place and we have the effect shown in figure 22.22 (a). This separation of colours one from another is dispersion and a simple way to produce it is to use the method of Practical activity 22.4. The spectrum obtained (figure 22.23) is described as impure only because the different colours merge into one another as in a rainbow and are not separated from one another. In fact the rainbow is a very good example of refraction and dispersion taking place at the same time with some T.I.R. as well. The difference is that in the prism the dispersed beam that leaves the first surface emerges into the air at the other refracting surface. In the case of the rainbow the prism is replaced by a tiny raindrop and, after the dispersion at the first refraction, the light then suffers reflection at the back surface of the water and then another refraction with more dispersion at emergence. Hence the rainbow! See how much of this is explained on the internet. The diagrams supplied there should be very clear and quite attractive.

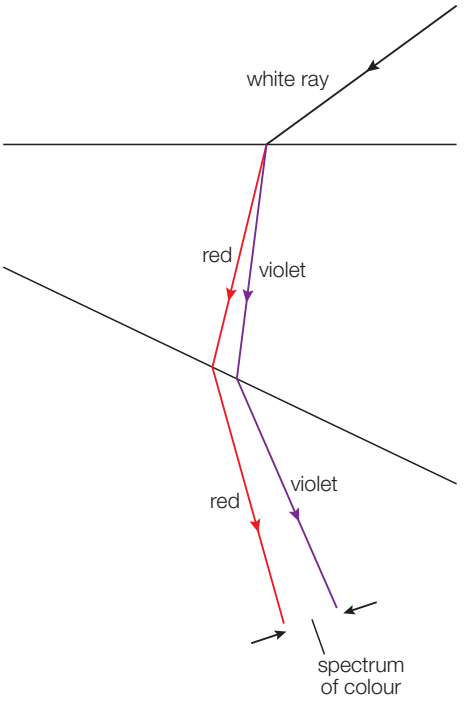


Figure 22.21 Dispersion of white light produced by two non-parallel refracting surfaces.

Note the difference between deviation and dispersion. Deviation is the change of course that ray undergoes on entering and leaving the prism. There are two deviations, one at each of the refracting faces, and the overall deviation is the sum of these. Dispersion is the separation of white light into its component colours, red, orange, yellow, green, blue indigo and violet.

In figure 22.22 (b), there is no refraction of the light since the angle of incidence is 0° and the waves of all frequencies take the same path through the rectangular block.

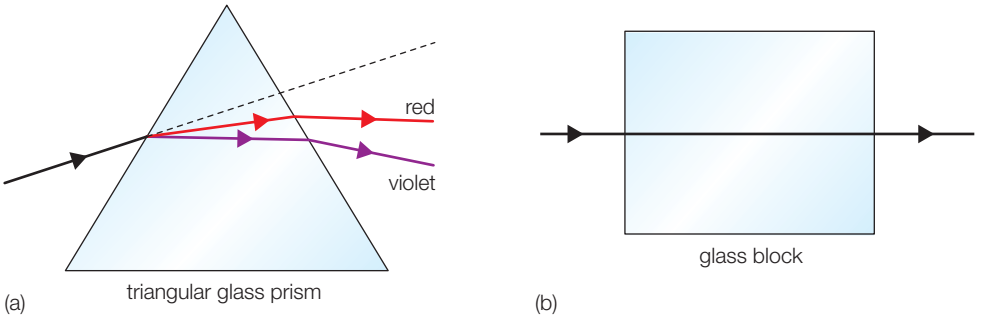


Figure 22.22 (a) Dispersion of white light by a prism. Light waves of different wavelengths (different colours) take differing paths because of different speeds in the glass. Faster colours are refracted less and slower ones more. (b)The angle of incidence on both faces of the rectangular glass block is zero, so there is no refraction and no dispersion of light into component colours and no deviation of the incident ray.

Practical activity 22.4

Producing an impure spectrum

You will need:

- prism with angles of 60°
- opaque screen with a small, narrow rectangular slit
- strong distant light source (strong sunlight is quite suitable for the experiment cannot be carried out in a dark room)
- white screen.

Method

1. Allow light from the source to pass through the vertical slit in the opaque sheet on to the prism (see figure 22.23).
2. Adjust the orientation of the prism so that the plane of incidence is perpendicular to the refracting surface of the prism.
3. Place the screen in position and adjust it so that it is roughly perpendicular to the direction of the emergent rays.

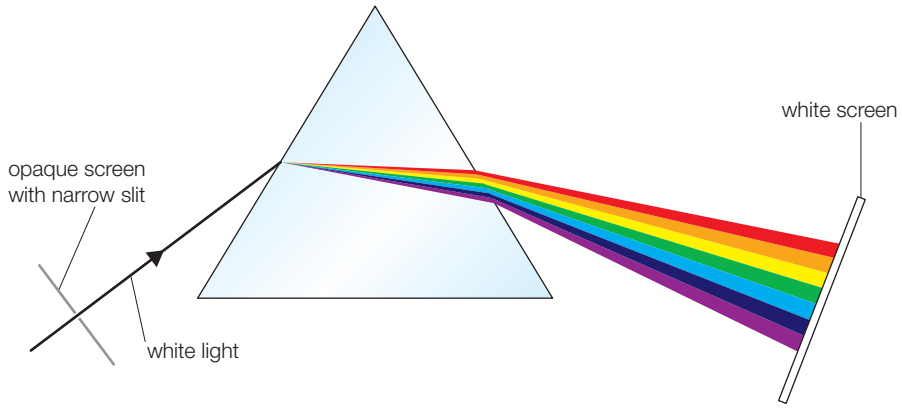


Figure 22.23 Producing an impure spectrum.

Observation

The spectrum will be seen on the screen. The colours will be blurred, overlapping with one another and so the spectrum is called 'impure'. To obtain a pure spectrum, we would have to use two converging lenses, in addition to the apparatus mentioned above, and the spectrum would appear as a series of colours strips next to one another.

ITQ4

With the help of the internet explain how mirages are formed:

- **in the desert;**
- **on the street on a hot day.**

ITQ5

Try this experiment: Put a coin into an empty vessel (preferably opaque) and position your head in such a way that the coin is just out of sight as shown in figure 22.24. Now slowly pour some water into the vessel, keeping your eye in the same position, until you just see the coin. Stop pouring. Draw a ray diagram to explain why the coin appears when there is sufficient water in the vessel but not before.

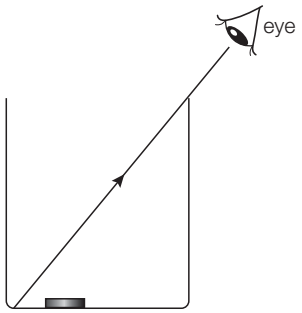


Figure 22.24

Newton and colour

Everyone knows that the colours with which we are so familiar today can all be found as components of what we know as 'white light'. It is very doubtful, however, whether this was known with absolute certainty by anyone before the late seventeenth century. People did observe, probably incidentally, that when sunlight passed through a prism the light emerging was coloured. They would have seen, also, and wondered about the colours of the rainbow and those who used telescopes would have noticed that images produced by them were often coloured. Whether they had any idea that the colours they saw were part of white light is very doubtful. Some thought that the colours seen were given to the white light by the glass; in other words it was thought that it was some property of the glass that gave rise to the colour of the light which emerged. We now know that it was the property of the light, rather than the glass, that caused light leaving the prism to be coloured and images formed by telescopes and other optical devices to be tinged with colour.

Newton would doubtless have used telescopes of one kind or another for looking at and studying the planets, rather as Galileo did before him. He

must have been irritated by the colours which tinged the images he saw – the effects of dispersion – caused by the lenses in the telescopes. He therefore set about studying the problem in the hope that he could discover the reason and try to avoid the problem. In fact, he did, to some extent, by replacing one of the telescope lenses by a curved mirror and a plane mirror in his 'Newtonian telescope'.

In his treatise on light, *Opticks*, published in 1704, he described an experiment that was very similar to the one we described in the last practical activity, except that it was not done in a laboratory, as we would have, but in his study. Not only was he able to separate the Sun's light into a wide spectrum and show that the colours all belonged to the white light from the Sun, but he was able to show, too, that one of these colours, having been 'isolated' and subjected to refraction through similar glass to that which produced it, came out unchanged. This proved beyond a doubt that the glass through which the light passed had no effect on the colour of the light which passed through it. But he did not stop there. He was able to show also that, by using a converging lens and another prism in the right way, he could produce the original white light he started with.

These experiments that Newton carried out provided proof that the colours we see and enjoy are basically mixtures of two or more of those comprising the spectrum. If an object appears green in white light, it is because the light it gives off has wavelengths belonging to the range for green found in the Sun's spectrum. It is not because the object changes the colour of the light from white to green.

Producing a pure spectrum

pure spectrum ›

Earlier, in figure 22.23, we saw that when white light passes through a triangular prism, a display of colours is seen. This display of the spectral colours from red to violet is called an impure spectrum because the colours merge into one another as in a rainbow. A '**pure spectrum**' in which the colours are better separated from one another may be obtained by using two lenses, the first one to render the incident light into a parallel beam and the other to focus the dispersed beam on to a screen. See figure 22.25. The first lens renders the light from the narrow slit parallel if the slit is placed in the focal plane of the lens. This parallel white light then falls on the prism shown in the figure. The light passing through this prism is dispersed and leaves the prism as a divergent dispersed beam in which all colours of the same wavelength travel in parallel directions. When this divergent dispersed light passes through the second lens, all rays of light of the same wavelength will be focussed at one point on the screen. Since a particular colour contains a range of wavelengths, there will be a range of points, all close together, where light of that particular colour will come to a focus. The result is that all the light of the same colour will be focussed in one band and if the slit in front of the lamp was horizontal, the image on the screen would consist of a set of horizontal bands of light ranging from red at the top through the colours present in the light emitted by the lamp ending in violet, if present. This spectrum is described as a pure spectrum. In this spectrum the individual colours are seen more distinctly (figure 22.25).

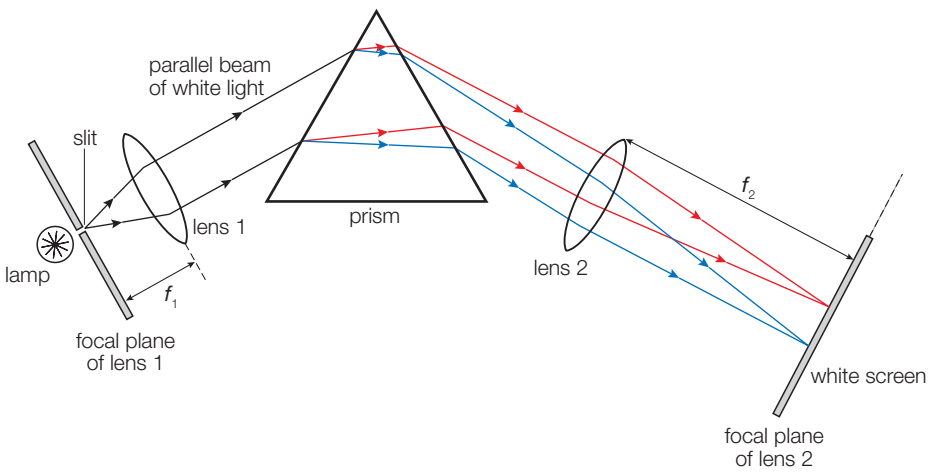


Figure 22.25 Producing a 'pure spectrum' of light.

Recombining the colours of the light spectrum

If the lens on the right in figure 22.25 is replaced by an inverted prism, close to the first, as shown in the diagram of figure 22.26, the spectral colours are recombined to form white light on the screen as shown in figure 22.27.

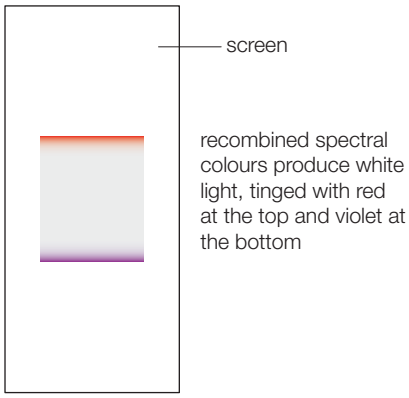


Figure 22.27 The colours obtained when lights are mixed.

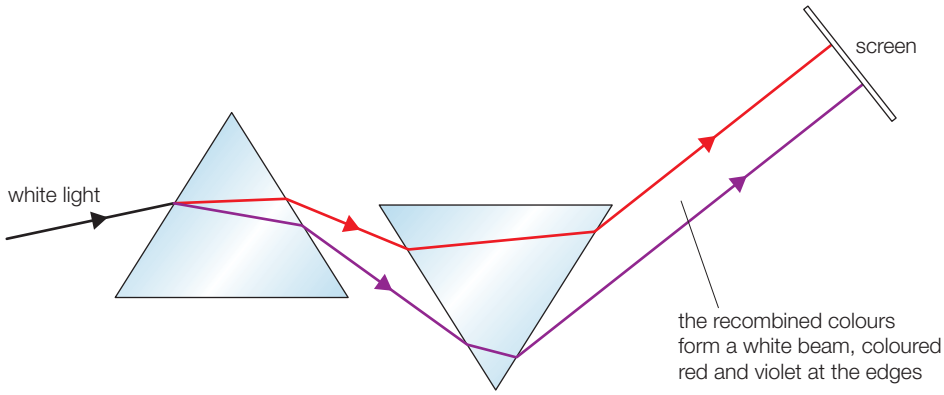


Figure 22.26 Recombining of spectral colours to produce white light.

The diagrams of figures 22.25, 22.26 and 22.27 suggest that white light consists of a spectrum of colours ranging from violet (of wavelength roughly 4×10^{-7} m) to red (of wavelength roughly 7×10^{-7} m).

When we look at light consisting of different colours we are really allowing light of different bands of wavelengths to fall on the retina and so what we perceive is the added effects of the bands of light radiation. Experiments have shown that there are three spectral colours, red, blue and green, called **primary colours** which, when taken two at a time, can produce three other colours, called **secondary colours**. Any required colour can be obtained by mixing two or more of these primary colours in the correct proportion.

primary colours ›
secondary colours ›

A primary colour of light cannot be produced by mixing other colours of light.

The diagram of figure 22.28 (a) shows the result of adding colours of light in a combination of coloured circles. Where circles of two different colours overlap, the resulting colour is that shown in the overlapping region. This colour in the overlapping region is the secondary colour. Thus where red and blue overlap we obtain 'magenta' (a secondary colour), and where blue and green overlap we obtain 'cyan' or blue green, another secondary colour. Yellow is the third secondary colour, obtained by adding red and green.

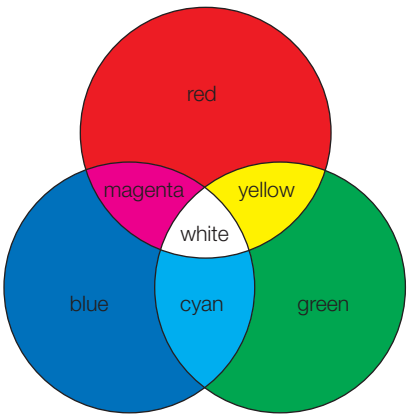


Figure 22.28 Addition of primary and secondary colours of light using the primary colour circles.

Where the three circles with primary colours overlap is the area common to all three circles. This area is white, showing that when red light, blue light and green light are added together, the result is white. So, (red light + blue light + green light) gives (white light), which as we know is colourless.

Applying a little bit of reasoning shows that magenta + yellow + cyan = white, since red + blue + green = white.

The Newton disc

The perception of 'white' colour can also be produced by spinning a disc with the three spectral colours, red, green and blue, each occupying a third of the disc area, at high speed. The disc appears white because each colour from the disc is retained in the eye for a fraction of a second. Because of a phenomenon known as 'persistence of vision', the overlapping impressions of red, blue and green within the eye gives the impression of 'white'. Such a disc is called a '**Newton disc**'.

Newton disc ›

The colours of objects

The colour an object is perceived to have is the colour of the light emitted or reflected by that object, and what this colour is depends on the colour of the light falling on the object.

Thus, a flower will appear red in white light if it absorbs all the other spectral colours apart from red. Again, if that flower is observed in magenta light it will again appear red, since the light falling on it is a mixture of red and blue and while the blue will be absorbed, the red will not absorbed, but reflected and so it will appear red. If, however, the flower is observed in cyan light, it will appear black, since the spectral colours falling on it, green and blue, are both absorbed by the flower and no light will be emitted by it. Since no light is reflected by the flower it will appear black.

Worked example 22.4

A yellow tie with blue dots is viewed in green light. What will be the perceived colour of the yellow background and the blue dots?

Solution

In white light, the yellow part of the tie will emit red and green light (which make yellow) and absorb blue. The blue dots will emit blue light and absorb red and green.

When illuminated with green light, green will be emitted by the yellow part of the tie, (since yellow = red + green), but absorbed by the blue dots. The tie will therefore appear green with black dots.

Worked example 22.5

What is the result of mixing red paint with blue paint?

Solution

The red paint absorbs (blue + green); the blue paint absorbs red.

So the red paint + the blue paint will absorb (red + blue + green), that is they will absorb white. The result is that the paint appears black.

Painters and artists know, however, that when mixing red paint with blue paint a purple colour results. This is because the paints used do reflect some red and some blue light. A similar effect is obtained when mixing yellow and blue paints. The result is green (not black) since both paints do reflect some green.

Chapter summary

- When a ray of light leaves one medium for another in which its speed is different, the ray bends and travels in a different direction (unless it strikes the interface between the two media normally). This bending of the ray is known as refraction.
- The amount of bending the ray undergoes depends on the ratio of the speeds of the light in the two media.
- There are three laws governing refraction:
  - The incident ray, the refracted ray and the normal at the point of incidence all lie in the same plane.
  - The refracted ray and the incident ray are on opposite sides of the normal.
  - For a given pair of media, the ratio of the sine of the angle of incidence and the sine of the angle of refraction is constant. This is known as Snell's law.
- The amount of refraction produced by a medium is related to the speed of light in that medium: the slower the speed of travel, the greater the refraction.

- The refracting ability of a medium is indicated by the refractive index of that medium.
- Refractive index of medium, $n = \frac{\text{speed of light in a vacuum}}{\text{speed of light in the medium}}$
- In practice, the speed of light in air is used in the definition rather than the speed in a vacuum.
- The refractive index also equals $\sin \theta_1 / \sin \theta_2$, where θ_1 and θ_2 are the angles with the normal at the interface made by the ray in air and in the medium, respectively.
- Two important equations relating to refraction of a boundary between two media, 1 and 2, are

$$\frac{c_1}{\sin \theta_1} = \frac{c_2}{\sin \theta_2}$$

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

  where c represents the speed of light in the medium, θ is the angle the ray makes with the normal in the medium, and n is the refractive index of the medium.
- A light ray is said to behave critically if its angle of incidence in one medium is such that the refracted ray travels along the interface of the two media. This angle of incidence is called the critical angle.
- For a ray to behave critically, the light must be moving from a slower to a faster medium, that is to a medium with a lower refractive index.
- The critical angle, θ_c, at any interface depends on the nature of each of the two media concerned, and is given by the formula:

$$\sin \theta_c = \frac{\text{value of the smaller refractive index}}{\text{value of the larger refractive index}}$$

- If the angle of incidence of a ray in the slower of two media at an interface between them is greater than the critical angle, the ray will not enter the faster medium, but will be reflected back into the slower medium. This effect is called total internal reflection. The reflected ray follows the laws of reflection.
- Waves of all the frequencies that make up white light travel at the same speed ($3.0 \times 10^8 \text{ m s}^{-1}$) in air, but not in any other medium.
- Of the colours of light that make up the visible spectrum, red light travels fastest. The slowest is violet.
- When white light strikes a transparent glass or plastic prism, because of the differences in speed between the different colours, light of each colour takes a different path. This separation of colours is called dispersion.
- The deviation shown by the dispersed beam of light varies with the colour. Red, with the fastest light waves, is deviated least and violet is the most deviated.
- Light of any of the colours of the visible range may be obtained by mixing two or more of the three basic colours red, blue and green (the primary colours).
- These three colours are therefore called the additive primary colours for colour mixing: 'additive', because they have to be added together to produce other colours in the visible range.
- When two primary colours of light are added, the result is a secondary colour. There are three secondary colours, namely magenta (red + blue), yellow (red + green) and cyan (green + blue).
- When the three primary colours of light are added, white (colourless) light is obtained.
- When light falls on an object, some of it might be absorbed and the rest reflected. The colour of the reflected light will determine the perceived colour of the object.
- When differently coloured paints are mixed, the colour of the mixture will be that of the light reflected by the mixture.
- The colour of an object will also depend on the colour of the light in which it is viewed.

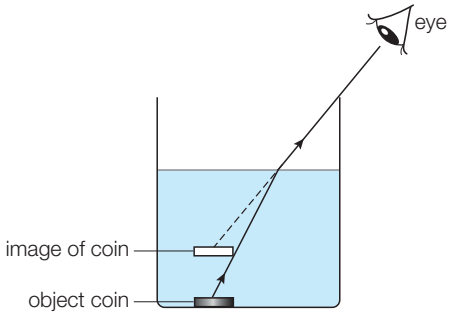


Figure 22.29

Answers to ITQs

ITQ1 The order of the speeds is c_w, c_o, c_a. The order of refractive indices is n_a, n_o, n_w. **(i)** 24.6°; **(ii)** 33.7°

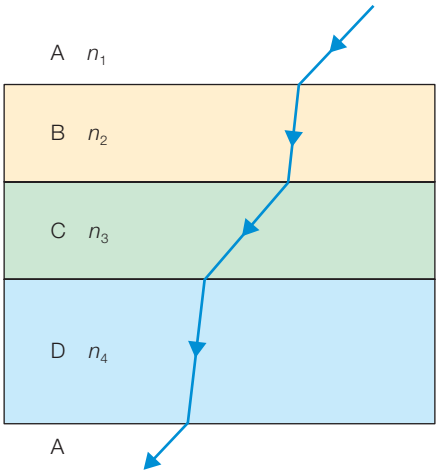
ITQ2 Using the formula $n_1 \sin\theta_1 = n_2 \sin\theta_2$, we have $1.0 \sin 0° = n \sin\theta$, where θ is the angle of refraction in the medium, so $n \sin\theta = 0$ and since $n_1 \neq 0$, then $\sin\theta = 0$ and $\theta = 0$. Thus, the angle of refraction = 0°.

ITQ3 Critical angles are 24.6° (for air); 33.7° (for water),

ITQ4 Internet provides the answer.

ITQ5 Rays from the coin are refracted away from the normal at the surface of the water, so that you can now see the coin. Its apparent position is higher than its real position. See figure 22.29.

Examination-style questions



1 A ray of light passes from a transparent medium, A, and then successively into transparent media B, C and D as shown in the diagram on the left. The refractive indices of the media A, B, C and D are respectively n_1, n_2, n_3 and n_4.

(i) Which of each of the following pairs of the media has the greater refractive index?

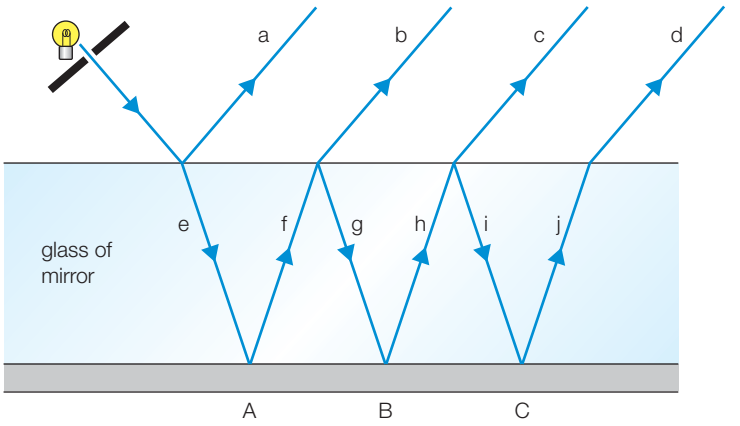
(a) A and B

(b) B and C

(c) C and D

(ii) Which pair(s) of the media listed in (i) have equal or roughly equal refractive indices?

2 A light ray strikes the surface of a thick glass mirror and undergoes multiple reflections at A, B and C.



(i) Which of the rays labelled a–j in the diagram are the result of:

(a) partial reflection?

(b) total internal reflection?

(c) refraction?

(ii) How many images of the bulb can an observer see when he looks into the mirror? Which of these will be the brightest?

(iii) One of the images seen may not be coloured. Which one?

(iv) Explain why the others are very likely to show tinges of colour.

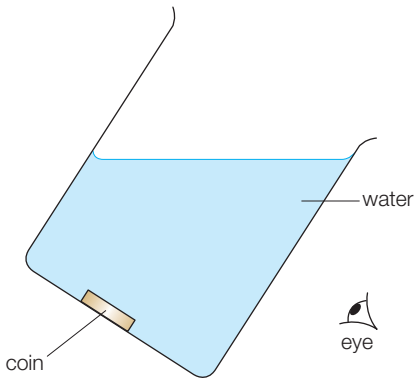
3 The values of the refractive index with air for four transparent substances are as follows:

glass	1.63	Perspex	1.53
plastic	1.40	water	1.33

(i) Arrange the values of the critical angle for the following interfaces in order of increasing size:
   (a) glass–Perspex
   (b) plastic–Perspex
   (c) water–Perspex

(ii) Calculate the value of the critical angle for each of these interfaces.

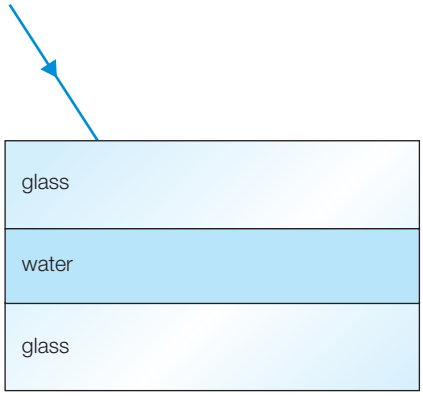
4 A beaker is half-filled with water and a small coin placed at the bottom of it. On leaning the beaker forward and looking upward towards the surface of the water, as shown in the diagram, an observer sees an image of the coin above the level of the water.



(i) Draw two rays from a point on the coin to show the formation of the image seen by the observer. Name the effect responsible for the formation of the image.

(ii) With the aid of a ray diagram, explain why a bunch of mangoes hanging over a swimming pool appears to be higher than it really is to someone under the water in the pool looking upwards at the mangoes.

5 A ray of light in air enters a combination of three rectangular parallel-sided 'blocks' in contact with one another as shown in the diagram. The refractive indices of glass, water and air are respectively 1.53, 1.33 and 1.00.



(i) What can you say about:
   (a) the path of the ray in the two blocks of glass?
   (b) the directions of the incident and emergent rays?

(ii) Name the law you have used to arrive at the answer to (i)(a).

(iii) Draw the path of the ray through the combination of blocks and out again into the air.

(iv) Indicate by the letter 'D' the deviation produced (if this applies) or the lateral displacement produced (if this applies).

6 The light of a certain discharge lamp consists of green and blue only. When viewed in this light, what would be the colour of:

(i) a dark blue suit?

(ii) a white suit?

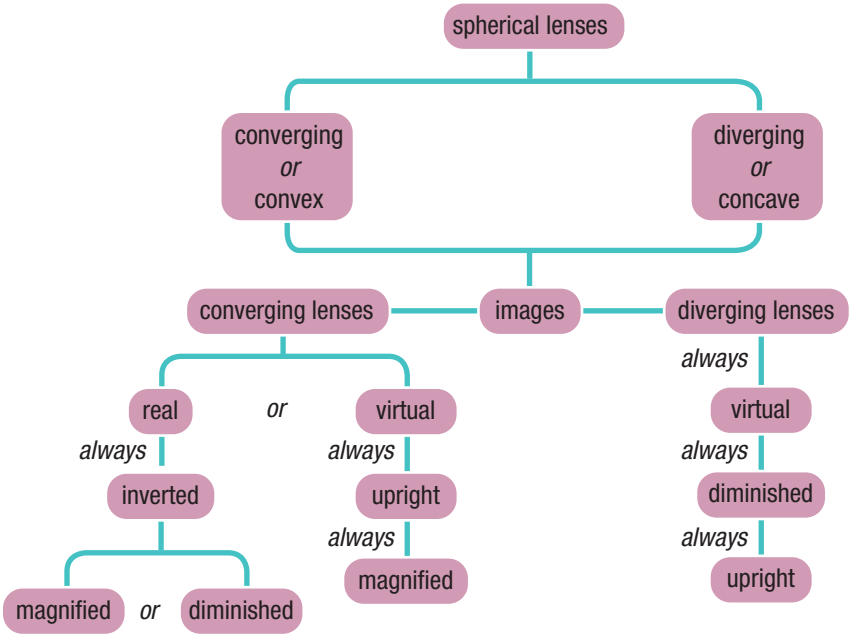
(iii) a yellow rose?

(iv) a pink dress?

23 Lenses

By the end of this chapter you should be able to:

- recall that lenses cause parallel light rays falling on them to either get closer together (or converge) or get further apart (or diverge), on passing through the lens
- recall that converging lenses cause the convergence of parallel light and diverging lenses cause the divergence of parallel light
- recall that converging and diverging lenses may be divided into sub-classes depending on the shapes of their faces
- recall that converging lenses are thicker in the centre than at the edges, but that diverging lenses are thinner at the centre than at the edges
- recall that both converging and diverging lenses may be considered to act like a series of truncated prisms of continuously decreasing and then increasing angle from one end of a diameter to the other
- draw ray diagrams showing the effect of (i) a thin converging lens on a narrow beam of parallel light rays travelling near to the axis, and (ii) a thin diverging lens on a narrow beam of parallel light rays travelling parallel to the axis
- recall the meanings of the following terms: principal axis; principal plane; principal focus; focal length; focal plane; paraxial ray; magnification
- describe and perform experiments to determine the focal length of a converging lens
- state the difference in nature between a real image and a virtual image
- perform experiments to locate real and virtual images
- use the relation magnification = $\frac{\text{image distance}}{\text{object distance}}$ to solve problems
- draw ray diagrams to show how images are formed by converging and diverging lenses
- use scale drawings to solve problems involving the formation of images by converging and diverging lenses
- use the 'real is positive' lens equation to solve problems involving the formation of images by converging and diverging lenses



Lenses: what are they and how do they work?

Lenses are devices made of transparent material such as glass or Perspex. Because of their shape, they will cause parallel light falling on them to either come to a point when they leave the lens (or converge), or spread out more and more (or diverges) when the light leaves the lens. All lenses, regardless of their shape, will have one or the other of these two actions on parallel rays and the word used to describe a particular lens will depend on this action on parallel rays. If the lens causes convergence of the parallel rays, it is called a converging lens. If it causes divergence of parallel rays, it is called a diverging lens. Thus the lenses used in correcting short-sightedness and long-sightedness would be diverging for the former and converging for the latter. Essentially, converging lenses reduce divergence in divergent light beams and increases convergence in convergent light beams. Divergent lenses have the opposite effect. They will increase divergence in divergent light beams and reduce convergence in convergent light beams.

The shapes of lenses

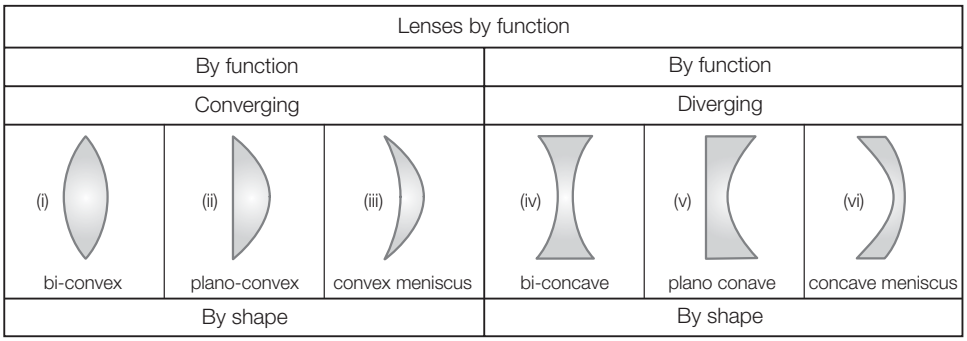


Figure 23.1 How lenses are described.

Lenses, in general, are either thicker in the centre than at the edges or thinner in the centre than at the edges. The most common shapes are shown in figure 23.1. You will notice that those lenses under the caption 'converging' are all thicker in the middle than at the edges and they can readily be identified as converging lenses. By contrast, those captioned 'diverging' are thinner in the middle than at the edges. This may be seen as a broad classification of lenses, one that is based on their 'contour'. Whereas the ones that are thicker in the middle bring about convergence in parallel rays, those that are thinner in the middle bring about divergence in parallel rays.

We can do a further sub-classification based on their shape. Lens (i) is called bi-convex ('bi' suggests two, in this case 'both') because both of its surfaces are convex (bulging outwards towards the observer on both surfaces), lens (ii) is called plano-convex for a reason that will be obvious and lens (iii) is described as convex meniscus, clearly because it has the shape of a liquid meniscus. The diverging lenses (iv) to (vi) show the opposite features and so will have the opposite effect on parallel rays to the lenses labelled (i) to (iii). The names used to describe them all fit their shapes.

Spherical lenses

At least one of the two refracting surfaces of lenses may have the shape of part of the surface of a sphere. Such lenses are called spherical lenses. 'Cylindrical' lenses have surfaces shaped like part of the curved surface of a cylinder. We will not be concerned with these. We are concerned here only with spherical lenses.

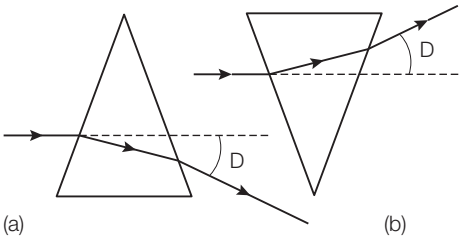


Figure 23.2 Deviation produced by the top half of (a) converging lenses and (b) diverging lenses.

How do lenses cause parallel light to converge or diverge?

To answer this question we must examine what happens to a ray of light when it strikes the surface of a lens. First, you must remember that a ray travelling parallel to the 'base' of a prism will be refracted towards the base of that prism. See figure 23.2.

Remember from chapter 22 that the deviation produced by a prism increases as the refracting angle of the prism increases. This is shown in figure 23.3 (a) Thus the truncated prisms into which the lens has been divided (as in figure 23.3 (b)) each produce a different deviation and so as the point of incidence of a ray gets closer to the centre of the lens the deviation of the incident ray becomes smaller and smaller. In the case of an actual lens the refracting angle of the truncated prisms decreases continuously as you get closer to the centre until, at the centre, the deviation becomes 0. A ray striking the lens at the centre of one of its surfaces therefore suffers no deviation on passing through the lens, since at this point the truncated prism has parallel faces. Refer back to figure 22.6 in chapter 22.

'Truncated' means 'with an end cut off'.

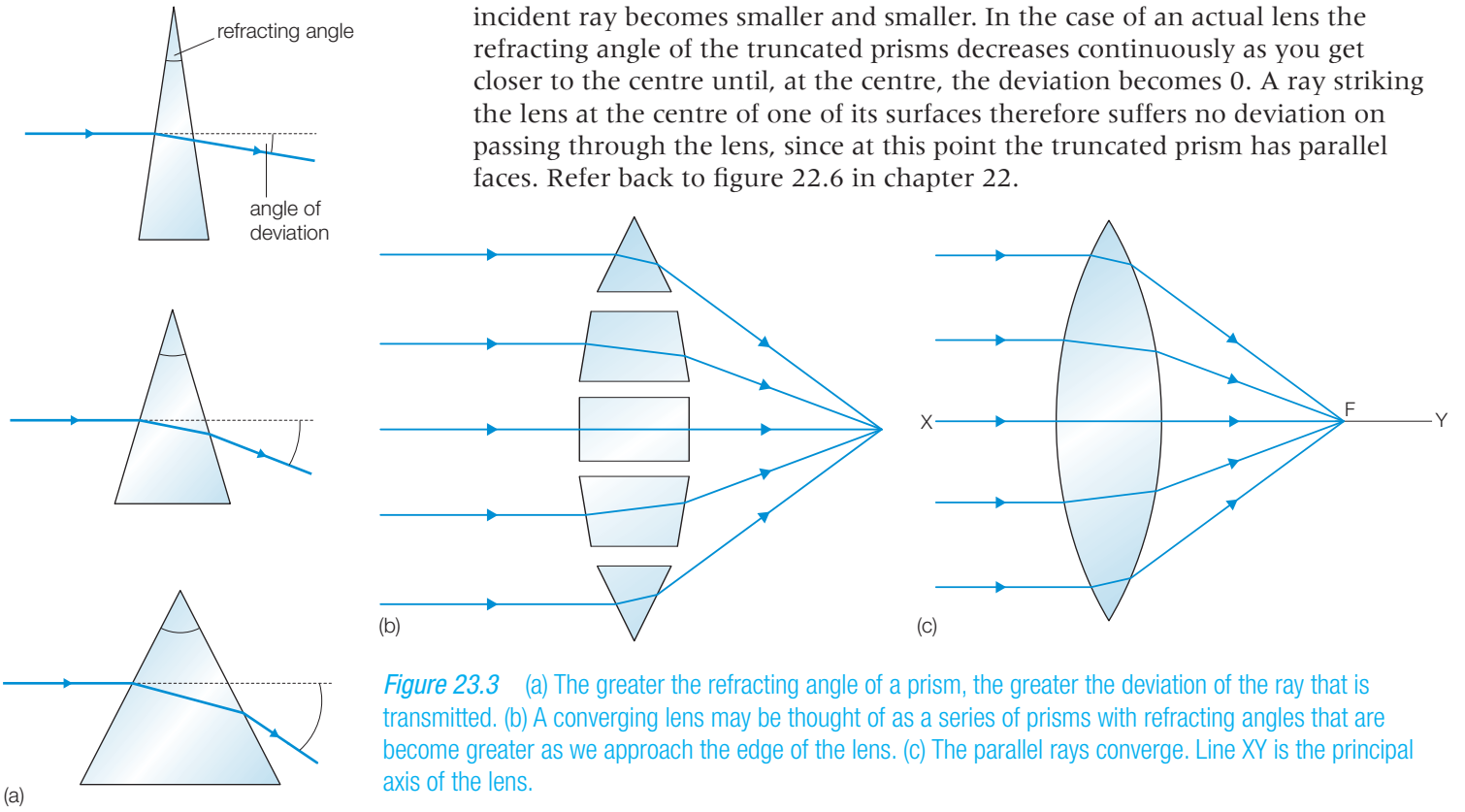


Figure 23.3 (a) The greater the refracting angle of a prism, the greater the deviation of the ray that is transmitted. (b) A converging lens may be thought of as a series of prisms with refracting angles that are become greater as we approach the edge of the lens. (c) The parallel rays converge. Line XY is the principal axis of the lens.

principal axis ›
principal focus ›

We call the imaginary line that coincides with the path of this undeviated ray XY the **principal axis** of the lens. In Figure 23.3 (c) the point F on XY shown is the **principal focus**. The result obtained when a beam of rays that is parallel to principal axis meets the lens is that they all come to a point on the other side of the lens. This point at which the emergent rays all meet is called the principal focus of the lens. It is marked F in figure 23.3 (c). This is a **real focus**, 'real' because the rays giving rise to it can meet on a screen and give an image on that screen.

real focus ›

A converging lens has a real focus.

In figure 23.4 (a)) the top half of the truncated prisms and the bottom half also are arranged with their vertices pointing towards the centre of the lens. We again have varying degrees of deviation of the emergent rays. This time the rays do not come together at all. Instead they do the opposite – they separate more and more away from one another as they get further and further from the lens – they diverge. Figure 23.4 (b) shows this. Again, as before with the converging lens, the ray that will pass through the lens undeviated is the one that passes through the centre of the lens where the refracting faces are parallel. Again the imaginary line that coincides with this ray is the principal axis of the diverging lens. The principal focus of the lens is the point behind the lens where these diverging rays *seem* to come from. To obtain this point we produce the emergent rays backwards behind the lens. This point is marked F′ in figure 23.4 (b). Immediately, we can see that such rays will give rise to a virtual image behind the lens and since there are really no rays where the focus is situated, the image is a virtual one and the focus a **virtual focus**, virtual here, as always in optics, meaning 'not real'. (Does the word 'virtual' have the same meaning in electronic gaming?) We say, then, that:

virtual focus ›

A diverging lens has a virtual focus.

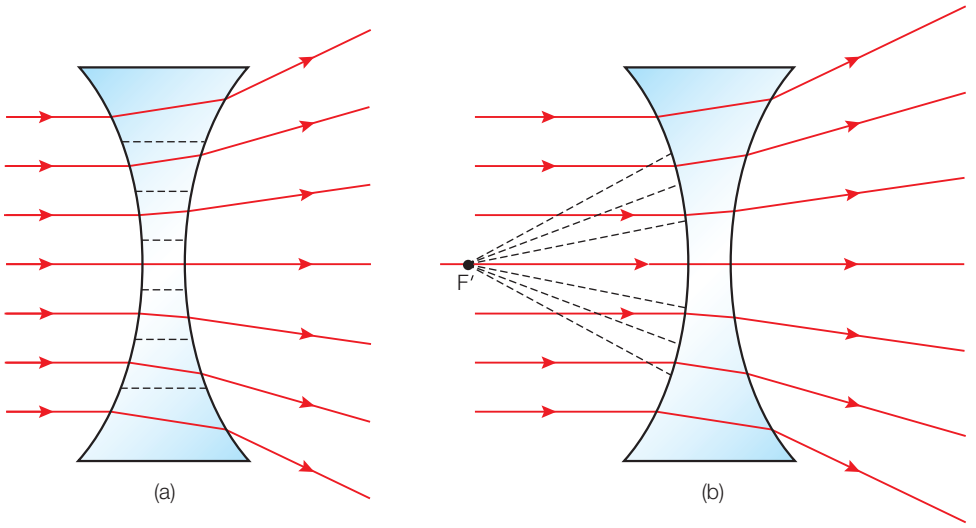


Figure 23.4 (a) Deviation of the incident rays depends on the size of the refracting angles of the truncated prisms – the larger the angle, the larger the deviation. (b) Parallel beams emerging from a concave lens have a virtual focus.

The first law of refraction tells us that, when refraction takes place, the emergent ray is in the same plane as the incident ray and the normal at the point of incidence. If, then, in Figure 23.3 the incident rays striking the lens and the normal were in the plane of the page, then the plane of the emergent rays will themselves be in the plane of the page and all the rays shown would be in the plane of the page. If these rays were now rotated through 180° about the principal axis of the lens, a cylindrical beam of rays each travelling parallel to the principal axis of the lens would emerge as a cone of rays on the other side of the lens, the vertex of this cone being the principal focus of the lens.

If we applied this reasoning to the diverging lens which is thinner in the middle than at the edge, a cone of rays would again be obtained, but truncated. The vertex of this cone would now be behind the lens. The base of the cone would be to the right of the lens and the 'virtual' vertex to the left. The virtual vertex would be the virtual focus of the diverging lens.

Lens terminology

The terms we use in discussing the action of lenses are many and it is very important that you understand not only the definitions of these terms, but also how we 'apply' the definitions in drawing ray diagrams.

The definitions we give below will apply to lenses that are 'equi-convex' and 'equi-concave', these terms meaning that both surfaces are convex and equally strongly curved in the first case and both concave and equally strongly curved in the other. In such cases, there will be a plane of symmetry through the edge of each type of lens.

We will give the definitions and in each case emphasise the significance and importance of the concept. Each of the concepts we will find to be extremely important in constructing ray diagrams showing the formation of images.

We begin with principal axis, principal plane and optical centre.

Principal axis

principal axis ›

We have already defined the **principal axis** of the lens as the imaginary line along which a ray must travel in order to pass through the lens undeviated. This line (labelled (i)) is clearly perpendicular to the lens if drawn from the centre of the lens surface (see figure 23.6). This definition would apply to both converging and diverging lenses.

The principal axis is the axis of symmetry of the lens system.

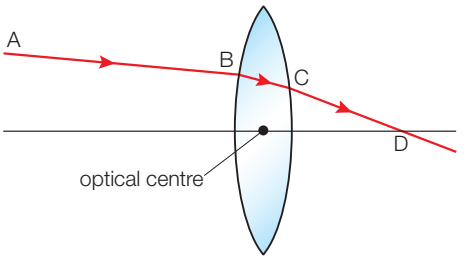


Figure 23.5

ITQ1

Justify the use of the word 'axis' in 'principal axis' by drawing the mirror image to the ray ABCD in figure 23.5. Explain the significance of the word 'axis' and how you used it to determine the answer.

Principal plane

principal plane ›

The **principal plane** of a lens (figure 23.6) is the plane of symmetry that passes through the edge of the lens. This plane of symmetry (labelled (2)) must clearly pass through the geometrical centre of the lens.

Its importance lies in the fact that in constructing ray diagrams rays are considered to undergo refraction at this plane instead of at the two lens surfaces.

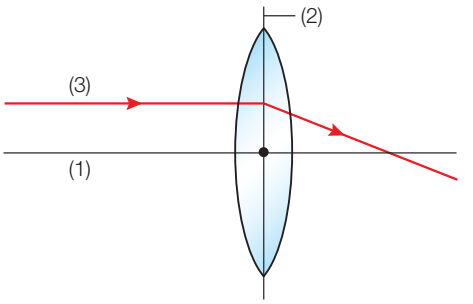


Figure 23.6 (1) Principal axis of a lens. (2) Principal plane of a lens. (3) Light ray being refracted at the principal plane.

The type of lens indicated in figure 23.6 is converging. Remembering that we are dealing only with thin lenses, the overall refracting action of the lens is represented as refraction at the principal plane only, rather than at both faces of the lens.

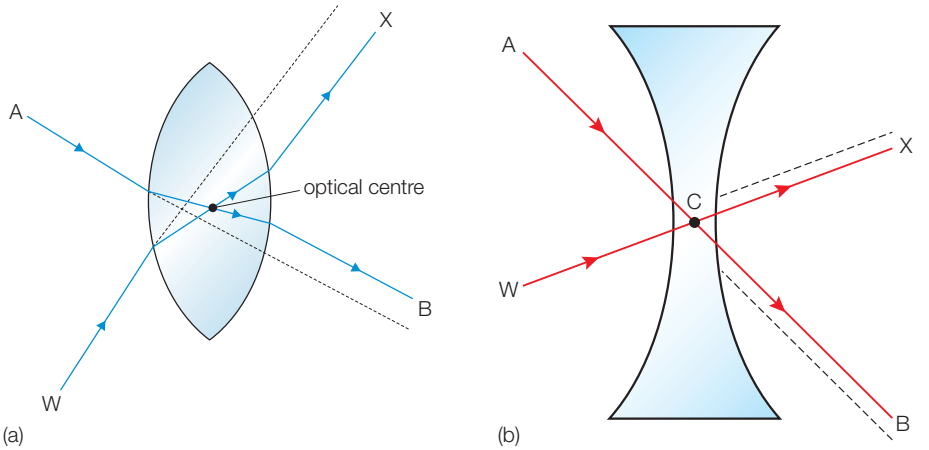


Figure 23.7 Because the lens is thick, the rays passing through the optical centre have some lateral displacement. These lateral displacements are small enough to be disregarded in practice.

Optical centre

optical centre ›

Figures 23.7 (a) and (b) show two rays, AB and WX, passing through a converging lens (in (a)) and the diverging lens (in (b)) without deviation, but with appreciable lateral displacement because the lens is thick. The point where these rays cross is the **optical centre** of the lens.

All other rays which pass through the lens without deviation will cross at this point. It is an interesting point to note that the surfaces where such rays strike the lens and leave it are parallel. (Remember the parallel-sided block of chapter 22?) The lenses we shall be dealing with will be considered to be very thin and so the lateral displacement produced will be small enough to be ignored.

The optical centre is one of the most important points and, as we shall see later, possibly the concept that is most used in drawing ray diagrams involving lenses. It is clear, from considerations of symmetry, that the optical centre will lie on the principal plane.

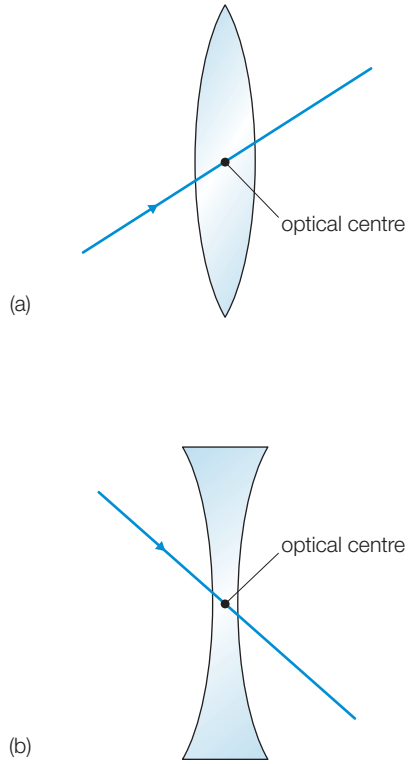


Figure 23.8 Rays passing through the optical centre of (a) a converging lens, and (b) a diverging lens. Note that there is no deviation or lateral displacement in either ray.

In drawing ray diagrams we disregard this small lateral displacement and we draw any ray passing through the optical centre as going straight through this point without deviation or lateral displacement. Such a ray is very important in drawing ray diagrams, sometimes being referred to as 'the **chief ray**'. Chief rays for a converging lens and a diverging lens are shown in figure 23.8.

chief ray ›

Rays passing through the optical centre of a lens, both converging and diverging, do so without deviation and without lateral displacement.

Principal focus, focal length and focal plane

principal focus ›

The **principal focus** of a lens is that point on the principal axis at which rays travelling parallel and near to the principal axis intersect (for a converging lens) or appear to have come from (for a diverging lens) after being refracted by the lens (figure 23.9). Clearly, since rays can fall on the lens from either side, there will be two foci for each type of lens, one on each side of the lens and the same distance from the optical centre in both cases.

Strictly speaking the rays used for finding the principal focus of a lens, both converging and diverging, should be, importantly, near to the axis if the lens. We must keep this point in mind when we carry out investigations using parallel rays.

We have met this before, but it is worth repeating that:

The principal focus of a converging lens is a real focus, since rays actually converge there. The principal focus of a diverging lens is virtual, since the rays diverge as they leave the lens and must be produced backwards to give the position of the image.

You are sure to come across the term 'focal point' used to represent the principal focus. This book discourages the use of the term, since there are millions of points which can be termed focal points, a focal point being any point where two or more rays are brought to a focus to produce an image. This point, properly described as the principal focus, is indeed the principal (meaning 'the most important' and a distinctive) point of focus of, particularly, rays that are:

- parallel to one another;
- parallel to the principal axis; and
- travelling near to the axis.

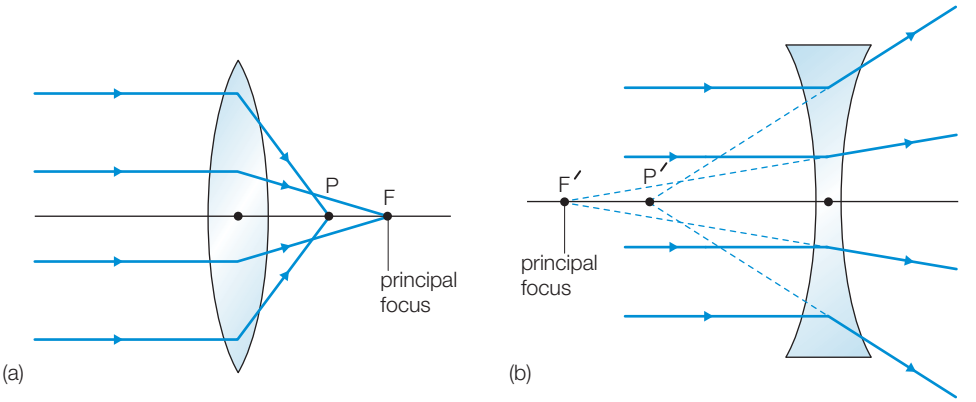


Figure 23.9 The principal focus of a lens is the point where rays travelling parallel and near to the principal axis intersect or appear to come from. (a) The principal focus of the converging lens is at F. It is not at P, since P is the point of focus of parallel rays that are far from the principal axis. (b) Similarly, for the diverging lens, the principal focus is at F′. It is not at P′, since the rays giving rise to that point are too far away from the axis. Rays travelling parallel and near to the principal axis appear to diverge from the point F′.

focal length

The **focal length** of a lens is the distance from the principal focus to the optical centre of the lens (figure 23.10 (a)). This distance is used in problems involving the action of lenses. It is, perhaps, the most important characteristic of a lens. The symbol used to represent it is '*f*'.

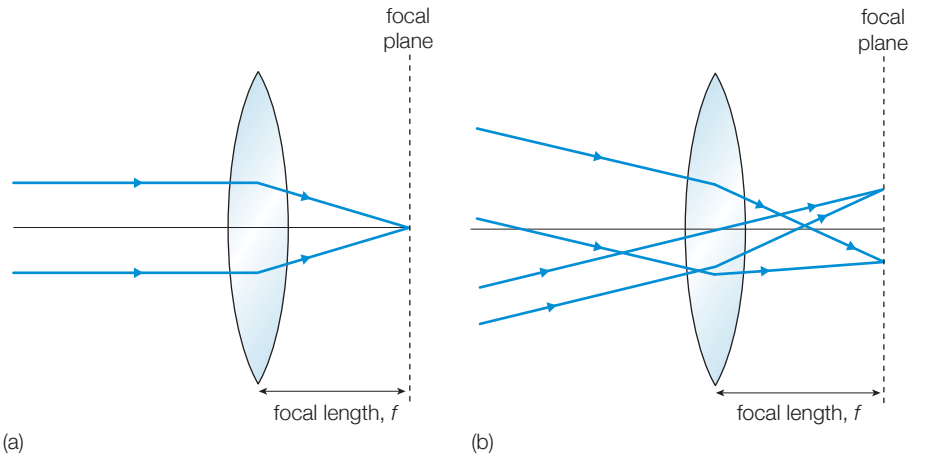


Figure 23.10 Focal length and focal plane. (a) Using rays parallel to the principal axis. (b) Using rays parallel to one another but not to the principal axis.

focal plane

The **focal plane** of a lens is the plane passing through the principal focus at right angles to the principal axis. The importance of this plane lies in the fact that all rays which:

(i) travel parallel to one another;

(ii) make a small angle with the principal axis; and

(iii) strike the lens near to this axis,

will come to a focus in this plane. Such rays are described as paraxial and are called **paraxial rays**. In order to determine where parallel rays will meet in a focal plane of a lens, a chief ray must be used. An example of the importance of this chief ray is shown in figure 23.11.

paraxial rays

In this figure you see two paraxial parallel rays a and b striking the lens. Since they are paraxial, they must meet in the focal plane of the lens. We cannot tell, however, what path they will each take on leaving the lens or

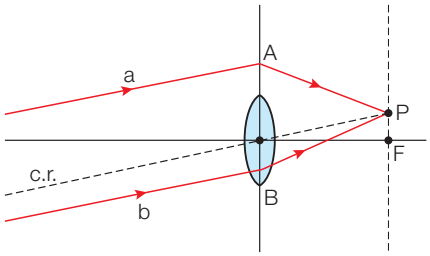


Figure 23.11 Use of the chief ray. Rays parallel to the chief ray (c.r.) will meet the focal place where the chief ray crosses it.

where in the focal plane they will come to a focus unless we use the chief ray. To do this, we draw the chief ray (marked c.r.) through the optical centre parallel to the given rays. Where this chief ray intersects the focal plane (at P) is the point at which all such (parallel and paraxial) rays should come to a focus. So we simply join the points of incidence, A and B, to this point P to show the paths of rays a and b.

If there were other rays travelling parallel to these two rays they would all meet at the point P in the focal plane. These rays must, however, have been paraxial.

When a chief ray is used in this manner, that is, to help to determine how a ray that strikes a lens will be refracted, it is described as a 'hypothetical chief ray'. This use of the chief ray applies to all lenses.

Worked example 23.1

A small light source is placed behind a cardboard screen that has a small hole. See figure 23.12. A narrow beam of light passes through this hole and falls on the lens paraxially at X. Use a chief ray to determine how the light beam will be refracted when it leaves the lens.

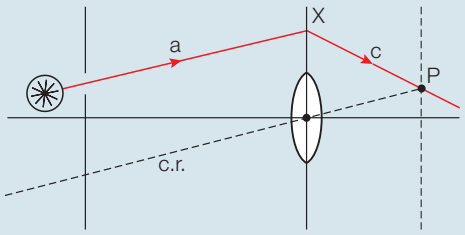


Figure 23.12 The importance of the chief ray. The chief ray shows how ray a will bend on leaving the lens. It will meet the focal plane at the same point as the chief ray.

Solution

The diagram of figure 23.12 shows the narrow beam denoted by 'a' reaching the principal plane of the lens at X. To find how it will be refracted by the lens we construct the hypothetical chief ray (c.r.) parallel to ray a to meet the focal plane at P.

Since ray a is parallel to c.r., they must both meet in the focal plane, and this means that ray a must bend as shown and meet c.r. at P. The path of ray a takes on leaving the lens is therefore shown by ray c.

Parallel rays from a point on a distant object such as the Sun, Moon, a distant street lamp or even a distant tree on a bright day would converge to a point in the focal plane of a converging lens when they are refracted by that converging lens. This fact can be used to find the focal length of a converging lens, as we shall soon see in Practical activity 23.1. Before that, however, there is one question:

Why are parallel rays associated with distant objects?

Consider a tiny lamp filament like that of a torch bulb (regarded as a point source) about 30 cm away from a converging lens. Consider, too, divergent light rays leaving the point source and falling on a small circular area of 1 cm diameter about the centre of one surface of the lens. The angle at the vertex of the cone of rays will be about 2° and the rays in this cone will certainly not be parallel.

If the source is moved ten times as far away, the angle will reduce to about 0.2°, and when 100 times as far away, the angle becomes about 0.02° and so on. It seems then that the angle at the vertex of the cone of the divergent beam (and also the angle between the rays in the cone generally) becomes smaller

and smaller as the source is moved further and further away from the lens. So for the rays to be parallel when they arrive at the lens, the source must very far away indeed, so far away that, even though the angle will always have a definite (though small) value, this value will be so very, very small that we might, for practical purposes, regard it as being zero. If, then, with the tiny light source very, very far away we allowed light from it to fall on the small area around the centre of the lens surface, the rays in the beam could be regarded as parallel and should come to a focus (a point) in the focal plane of the lens. This could be the basis for finding the focal length of a converging lens directly. We could use the Sun as a distant object by day, or the moon by night, or a distant street lamp by night , or even a distant tree on a sunny day when a lot of sunlight is falling on it to make it bright enough to give a bright image. This, then, will be the principle for a practical activity that we will carry out shortly. Before this, however, let us make clear that:

Parallel light rays are considered to come only from sources that are very far away, the further away the source, the more nearly parallel will be the rays coming from a point on that source.

Of course, we could ensure that rays reaching a lens from a source are paraxial by making the object distance large or, if this is not possible, by restricting the area over which light rays fall on the lens. By doing this we would be regulating the 'aperture' of the lens, the aperture being the area of the surface of the lens through which light will be allowed to pass through the lens. We will now see how the aperture of a lens affects the quality of the image the lens produces.

Practical activity 23.1

The effects of aperture size on the quality of the image produced by a lens

Aim

To investigate how the quality of an image produced by a converging lens is affected by the area of the lens surface exposed to the light from the source.

You will need:

- converging lens
- metre rule
- 6 volt torchlight bulb mounted in a lamp-holder connected to a suitable voltage source
- A lens-holder for the lens or, if not available, a lump of plasticine
- sheet of white paper stuck to a wooden surface provided with a base to serve as a screen
- three paper rings of outer diameter that of the lens, with central holes of diameter 1 cm, 3 cm, 5 cm respectively
- adhesive tape.

Method

See figure 23.13.

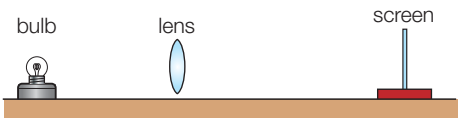


Figure 23.13 Examining the sharpness of images of a torch bulb.

The activity will be performed in two parts. Part (i) will examine the effect of the size of the central hole on the sharpness of the image produced and, having discovered which hole size gives the best, i.e. sharpest, images, use that hole size in part (ii) to 'discover' the relationship between the object distance and the image distance when the lens forms a real image of an object.

Part (i)

Aim

To investigate how the size of the hole through which light is admitted on to a lens affects the sharpness of the image of a distant object produced by the lens.

Method

1 Look for a dark area in your laboratory (perhaps the far corner of the room) or a dark room, if you have one. Set up the lens in the lens-holder (or on the lump of plasticine) on your laboratory bench with its plane vertical.

2 Adjust the height of the car bulb on the stand so that it is on the same level as the bulb. How would you do this? How would check that they are on the same level?

3 Switch on the bulb and place the screen in line with the bulb and the lens so that the three are in line. Move the screen to and fro along this line until an image is seen on the screen.

4 Stop moving the screen when you think you have the best-defined (or sharpest) image over the smallest range of distance. Measure and record on a table like the one shown below:

   (a) the diameter of the lens surface over which light is allowed to fall (or the lens aperture)

   (b) the quality (sharpness) of the image.

5 Now tape the ring with the largest hole symmetrically over the lens and repeat steps 2, 3 and 4.

6 Repeat steps 2, 3 and 4 for the remaining rings and summarise your results in table 23.1.

Table 23.1

Diameter of hole admitting light to lens, *d*/mm	Quality of image obtained on the screen (good/ satisfactory/poor)
Full area of lens	
5 cm	
3 cm	
1 cm	

Observations and comments

Observe and comment on aspects related to the image, e.g. sharpness, brightness and the range over which an image remained sharp. Draw inferences from your results concerning the effect of the size of the area of the lens exposed to light on the sharpness of the image.

Inference

Which of the rings gave the sharpest image of the filament over the shortest range of distance?

Draw a conclusion, remembering that this conclusion must be related to the aim.

Conclusion

Was there any relationship between the aperture (hole) size and the sharpness (or quality) of the image?

Part (ii)

Aim

To investigate the relationship between u and v for a converging lens where u = distance of object from the lens or the object distance and v = distance of image from the lens or the image distance.

Method

Having discovered which hole size gives the best image in terms of sharpness and brightness, follow these steps:

1 Decide on a range of object distances which you will use. This range should be as wide as possible once the image can be clearly seen.

2 Divide up this range of distances into six even and convenient intervals.

3 Draw up a table like table 23.2 to record your readings.

4 Use the object distances decided upon and obtain two 'independent' image distances for each object distance you decide to use. Independent image distances are distances taken one after the other. This means that you obtain one 'best value' of the image distance, and then obtain a fresh value having made a fresh start for that determination. Note that each of u and v is measured from the optical centre of the lens. How will you carry out this measurement if the optical centre is inaccessible? See figure 23.14 for a hint.

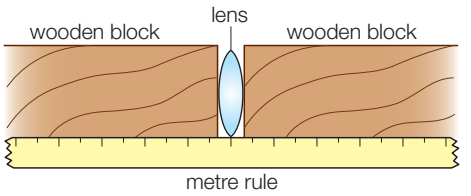


Figure 23.14 Measuring the thickness of a lens. From this measurement you can determine the position of the optical centre.

5 Record these values of u and v in table 23.2. How many significant figures will be permissible in recording these measurements? Discuss this together and then decide.

6 Repeat the procedure at 4 and 5 for each of the remaining values chosen for u.

Table 23.2

Object distance, u/cm	First value of image distance, v/cm	Second value of image distance, v/cm	mean value of image distance, v_m/cm	Reciprocal of v_m, $1/v_m$	Reciprocal of u, $1/u$	$(1/u) + (1/v)$
1						
2						
3						
4						
5						
6						

Use of the readings obtained

MATHEMATICS: equation of a straight line

Since you would have noticed that the values of the image distances changed as the values of the object distances changed, first plot a graph of v_m against u. What is the shape of this graph?

Can you draw any firm conclusions from it as to the way in which u and v are related?

Plot, also, the graph of $1/v_m$ against $1/u$.

Whichever of the two graphs gives a straight line, this will be the graph to use to draw the conclusion as to the relation between u and v. Having obtained a straight line graph for one or the other of the two graphs, now write the algebraic equation that represents the graph. This equation will represent the relationship between u and v for the lens. This will be all that can be said concerning u and v for the time being.

You should, however, be able to predict which of the two graphs is more likely to be linear (and therefore more useful to the arrival of a conclusion) from the way in which corresponding values of u and v_m behave.

Having plotted the two graphs, now add the reciprocals of u and v in the last column of the table. Have you noticed anything significant about the totals? If you have, do what you think is necessary to be able to draw a conclusion from the activity.

We can put to use our knowledge of the definition of the focal length to find the focal length of a converging lens using a distant object. In order to do this we must use parallel rays which are paraxial. We can arrange for the rays striking the lens to be:

(i) parallel by using a bright distant object (e.g. the sun, a distant tree, a distant building); and

(ii) paraxial by using a cardboard ring (called a 'lens stop') as in the last activity over the lens to reduce the area over which light is received by the lens and so the angle made by the rays with the lens axis.

Note: if the sun will be used as the object, a stop will not be necessary (why not?). If, however, there is no direct light available from the sun and a distant tree or building is used we must be sure that all the rays striking the lens are very nearly parallel although they will be coming from an object that is not at infinity like the sun or the moon.

The way to ensure this is to use a lens-stop to make the aperture of the lens small. But then doing this will reduce the amount of light forming the image, which will make for an image that is not bright enough to judge its quality. We must therefore use our judgment and decide what are the conditions that will give us 'the best of both worlds' – a bright image that is sharp with paraxial rays.

Practical activity 23.2

Find the focal length of a converging lens using a distant object

You will need:

- converging lens
- white screen
- half-metre rule
- sunshine directly from the Sun or, if there is no Sun, a distant building or a tree.

Method

Caution! Take care never to look directly at the Sun even if you are wearing sunglasses. What the lens you are using will do to parallel rays from the Sun is exactly what your eye-lens can do to them, namely, to focus these intense rays on your retina with disastrous effects.

1 Hold the lens so that its axis points in the direction of the Sun.

2 Place the screen behind the lens to receive light rays focussed by the lens.

3 Move the lens and/or the screen along the axis of the lens until the sharpest image of the Sun (or the tree or building) is obtained on the screen as a tiny, brilliant spot of light (figure 23.15) if the Sun is used, or a tiny tree or building, as the case may be.

4 Measure the distance between the centre of the lens and the screen at this point. Do not forget that the lens does have some thickness!

5 Repeat the procedure from the beginning four more times.

6 Calculate the mean of the five readings taken.

Observation

What are your observations regarding the orientation of the images of the tree or the building? Could you tell whether the image of the Sun is inverted or not? Are your images real? Why? What about their size? Try to explain the reasons for the observations you make.

Inference

The bright spot on the screen is the image of the Sun (or the tiny image of the tree or the building) formed by the parallel rays coming from the object being focussed there. This spot of light (or tiny image) is at the focus of the lens

Conclusion

The conclusion must be concerned with the focal length of the lens used.

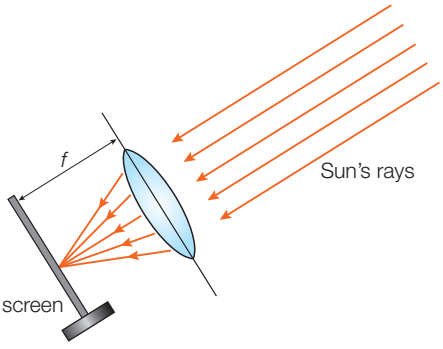


Figure 23.15 Finding the focal length of a converging lens using the Sun's rays.

How to locate (find the position of) a virtual image

The real image formed by a converging lens of an illuminated slit (a self-luminous object) can be easily located on a white screen, since the rays that form the image will come to a focus on the screen. How does one locate the position of a virtual object if there are no light rays present where the image is situated?

You may remember we used the method of no-parallax to locate the virtual image formed by a plane mirror in chapter 21. We can use a similar technique to find the position of a virtual image. Figure 23.16 shows the arrangement.

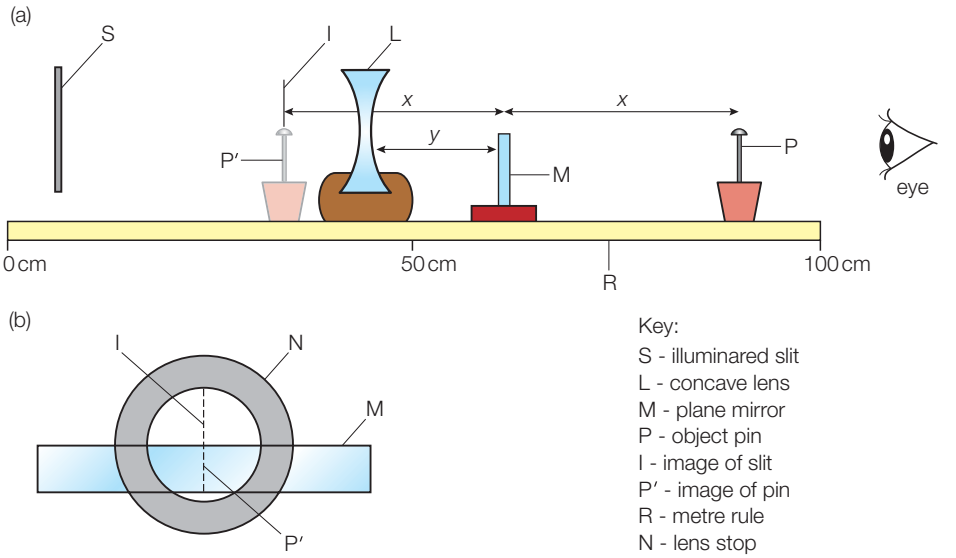


Figure 23.16 Locating a virtual image using a plane mirror. (a) Side view, (b) front view.

The concave lens, L, forms a virtual image, I, of an illuminated slit (or an optical pin) S. The image, P´, of the pin, P, formed by the plane mirror, M, is made to coincide with this image, I, of the slit by no parallax. The correct position of the plane mirror image, P´, is obtained by moving pin P to and fro and checking for no-parallax until it is obtained (see Practical activity 21.3 on page 311).Once no-parallax is obtained we measure the distances x and y. The distance from the image of the slit from the lens is $(x - y)$.

Practical activity 23.3

Find the focal length of a concave lens

You will need:

- concave lens
- 1 metre rule
- piece of card with slit cut out
- strip mirror (preferably a metal mirror)
- optical pin and cork.

Method

1 Position the concave lens, L, with its 'stop' and the illuminated slit, S, on a metre rule, with the lens at the 50 cm mark on the rule and the slit at the 0 cm mark. See figure 23.16 (a).

2 The image of the slit will be somewhere between the slit and the lens.

3 Place a plane strip mirror (preferably a metal mirror) across the rule (using plasticine) with the reflecting surface turned away from the lens and in front of the lens in such a way that the upper half of the lens is clearly visible above it. See figure 23.16 (b). When you look into the mirror you should see the image of the slit in the upper half of

the lens, but nothing (except your own image, of course) in the plane mirror.

4 Now place an optical pin, stuck in a cork, on the rule between yourself and the mirror such that you can see an image of the pin in the mirror. Adjust the position of the pin on the rule so that the image, P′, of the pin and the image, I, of the slit are together (one above the other) in the middle of the lens. See figure 23.16 (b).

5 With one eye closed, move your head to the left and to the right and look carefully at the two images, namely, that of the pin formed by the mirror and that of the slit formed by the lens. The chances are that, as you move your head, the two images will separate or show parallax.

The method would be exactly the same if you wanted to locate the virtual image formed by a converging lens.

6 Move the pin along the rule, having first checked as in 3, until you find that as you move your head from side to side the two images move together. When they do, you will know that they are showing no parallax (no separation) and this tells you that they are coincident. The image of the pin is in the same position as the image of the slit.

7 Measure from the rule the distance, x, between the mirror and the pin. This will also be the distance between the mirror and the image of the pin P. Measure also the distance, y, between the mirror and the lens.

8 Take the difference between x and y. You can see from figure 23.16 (a) that the distance of the image, I, from the lens is $(x - y)$. You have located the virtual image of the slit, since you know how far it is behind the lens.

We can now consider general cases of image formation of linear objects placed anywhere near to the axis of a lens, whether converging or diverging. We assume the objects to be near to the axis of the lens so that the rays used to form the image may all be paraxial that is to say, they are either parallel to the axis or they make a small angle with the principal axis.

Constructing ray diagrams

To show how images of objects are produced by lenses, we construct ray diagrams. These show how rays of light that leave points on the object are refracted by the lens. The rays striking the lens will either converge to a focus on leaving the lens, giving a real image, or diverge, giving a virtual image. If the image is real, it will be on the other side of the lens to the object. If virtual, it will be on the same side as the object.

In constructing ray diagrams, certain assumptions are made regarding the rays used:

- The rays used are all paraxial. This means that they make very small angles with the axis of the lens, and
- They strike the lens at points near to it centre.
- Rays coming from points very far from the lens are parallel to one another. Such rays will therefore converge somewhere in the focal plane since they are all paraxial.

Three types of ray are particularly useful when drawing ray diagrams:

1 A paraxial ray travelling parallel to the principal axis. Such a ray will pass through the principal focus after refraction (shown as ray 1 in the ray diagrams).

2 A paraxial ray that passes through the principal focus of a lens before striking the lens will emerge parallel to the principal axis, from the reversibility principle. (These rays are shown as ray 2 in the ray diagrams.)

3 A chief ray that is paraxial will pass through the optical centre of the lens without lateral displacement or deviation (ray 3 in the ray diagrams).

These three rays are referred to as 'rays 1, 2 and 3'. When we draw ray diagrams we use solid lines to show the paths of actual light rays and broken lines for construction lines.

First steps

The first steps to be taken are:

1 Draw a line to represent the principal axis across the page.
2 Draw in the principal plane of the lens as a line at right angles to the principal axis.
3 These lines will intersect at the optical centre of the lens. Mark this point C.
4 At the optical centre draw a tiny lens to indicate the type of lens being used, converging or diverging. There is no need for a lens to be drawn full size.
5 The two principal foci, F_1 and F_2, are marked on the diagram, equidistant from the centre.

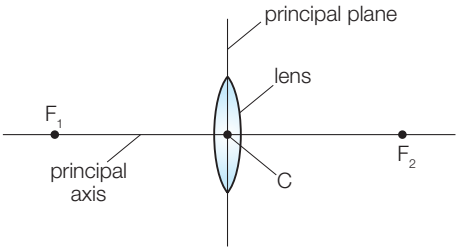


Figure 23.17 First steps in constructing a ray diagram.

The object is normally represented by an upright arrow.

Worked example 23.2

Construct a ray diagram showing how a converging lens forms an image of an object placed not far 'outside' of the principal focus of the lens, 'outside' the focus meaning beyond the focus away from the lens.

Solution

The steps are (see figure 23.17):

(i) Carry out steps 1 to 5 above.
(ii) Take two rays from the tip of the object up to the principal plane of the lens and let them be refracted by the lens in the way they are expected to do.

Referring to figure 23.18, we start by drawing a small arrow, OB, to represent the object at right angles to the principal axis with O on the axis and B above (or below, if preferred) the axis. We are now ready to do the construction. The position of the image of B will be that point where two rays that left the point B meet after being refracted by the lens. We begin the construction by drawing ray 1 from B parallel to the principal axis and refracted by the lens through the principal focus F_2.

(iii) We need a second ray from B to find the image of B. We now have a choice between rays 2 and 3 as the other ray to define the image of B. We will use ray 3.
(iv) Draw ray 3, a chief ray, from B straight through C, meeting ray 1 at M. M is therefore the image of B.

Alternatively, we could have used ray 2 as the second defining ray for the image of B. If we had done so, the ray from B would have been passed through the focus F_1 and, on meeting the principal plane would have left the lens parallel to the principal axis (reversibility principle!). A careful construction should have put B in the very same position as it had before. It really should not matter therefore which ray is chosen, 2 or 3, to give the image point, M, of the object point, O.

ITQ2
Give one reason why the image of an object on the principal axis of a lens must also be on the principal axis of the lens.

ITQ3
You are aware that the image of an object that is formed by a converging lens is inverted about a horizontal plane through the principal axis. Put differently we might say that 'what looks up on the object will look down in the image'. Explain why 'what looks left in the object must look right in the image', suggesting that there is also inversion about a vertical plane taken through the principal axis.

At this point we could, if we wished, carry out a construction to find the image of O. We generally do not do a construction, but assume that the image of O will be perpendicular to the principal axis since the object was perpendicular to it. We therefore put I, the image of O, vertically above M, since the image must be along the principal axis. Can you see why the image of O must also be on the axis? You would have noticed that the object just outside the focus on the left side of the lens produced an image far away from the focus on the other side of the lens.

The image of an object placed outside of the focus of a converging lens is:

- inverted about a horizontal plane through the principal axis;
- real, since light rays actually cross to form the image;
- is larger than the object.

In other words, the converging lens produces an inversion in a real image about both a vertical plane and also a horizontal plane, both planes passing through the principal axis. For example see figure 23.18 (b).

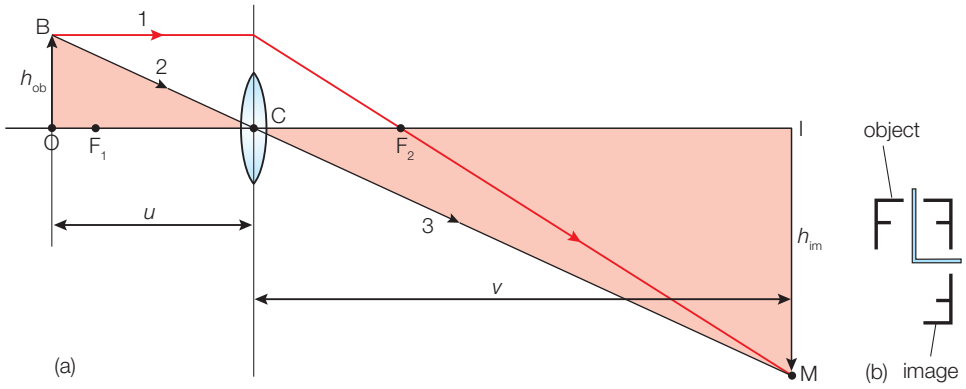


Figure 23.18 (a) Constructing a ray diagram. (b) Image twice inverted.

Worked example 23.3

Suppose the image IM in Worked example 23.2 were made the object, show that the image would be the object, OB.

Solution

The reversibility principle tells us that if, in figure 23.18, and starting with ray 1 we reversed the rays chosen as 1 and 3, they would follow the same paths but in the opposite directions. Ray 1 and ray 3 would meet at B, making B the image of M and the image of I would be O. It seems then that object and image can exchange positions. This is an important point to notice. It occurs, however, only when real objects produce real images. It does not happen when real objects produce virtual images, which we will show in Worked example 23.4!

ITQ4
Find out by doing a ray diagram whether the image produced by a converging lens of an object far away from the principal focus is larger than the object.

Thus so far we have seen that the image of an object placed outside of the focus of a converging lens is:

- inverted about both a vertical plane and a horizontal plane through the principal axis;
- real, since light rays actually cross to form the image; and
- is larger than the object.

What is the image like and where is it formed if the object is placed between the lens and the principal focus?

If you answered ITQ4, you would have discovered that the image was smaller than the object. It would appear then that somewhere along the axis the diminished or reduced image changes to an enlarged or a magnified image as the object moves along the axis. There must, therefore, be some point along the axis where they have the same size. Where does this happen? In light of the last feature mentioned, where must the object be placed for it and the image to have the same size? The answer is: at a point on the axis which is twice the focal length away from the lens. This point is designated X_1 on one side of the lens and X_2 on the other. We will prove this later when we describe an activity to investigate this.

Worked example 23.4

Draw a ray diagram to show the image formed by a converging lens of an object placed between the principal focus, F′, and the optical centre, C of the lens.

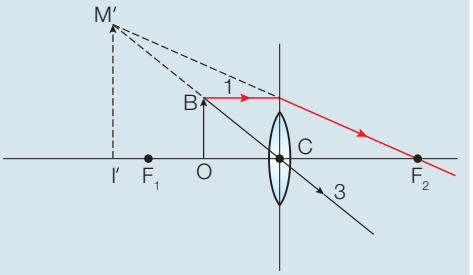


Figure 23.19 A converging lens producing a virtual image.

Solution

The procedure is exactly the same as for Worked example 232. However, when we construct rays 1 and 3, we find that they do not converge on leaving the lens, but diverge instead. Clearly from what has been stressed before in chapter 19 about rays having an optical device diverging and producing a virtual image, the image here has to be virtual, as is clear from figure 23.19. There the rays emerging from the lens are produced backwards to meet at M′. We presume, as before, that the image of O will be again on the axis, and so will be at I′.

So the image of an object placed between F and C for a converging lens is magnified, virtual and upright. Remember that the last two features always go together.

All virtual images are upright; all real images are inverted.

Construction of ray diagrams for diverging lenses

You will be surprised to learn, but should be relieved to know, that it is much easier to study the action of diverging lenses in forming images than it is to do so for converging lenses. Not only is the actual construction straightforward, but, happily, the kind (nature) of image obtained is always the same, regardless of the position of the object on the lens axis. In figure 23.20, ray 1 is drawn from B parallel to the axis and on refraction must *seem* to have come from principal focus, F_2, by definition. Hence the behaviour of ray 1. The behaviour of ray 3 should not be of concern. It is the chief ray and will pass straight through the optical centre, as expected. The divergent bundle of rays arriving at the lens from B leave the lens diverging even more than before (surprised?). The rays emerging from the lens must, as usual, be produced backwards to give the position of the virtual image obtained.

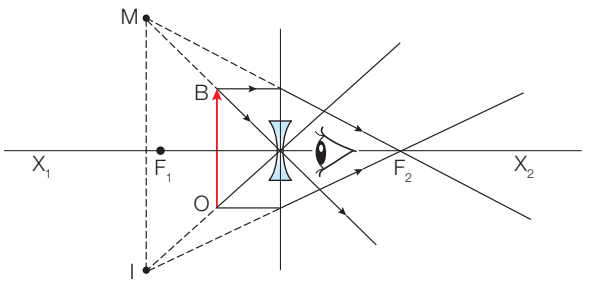


Figure 23.20 How a concave lens forms a virtual image.

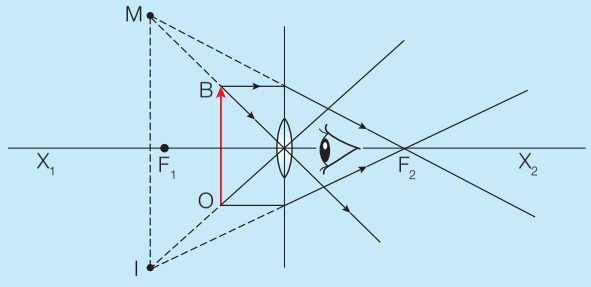
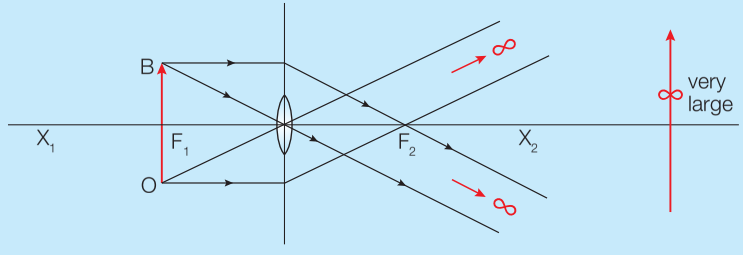
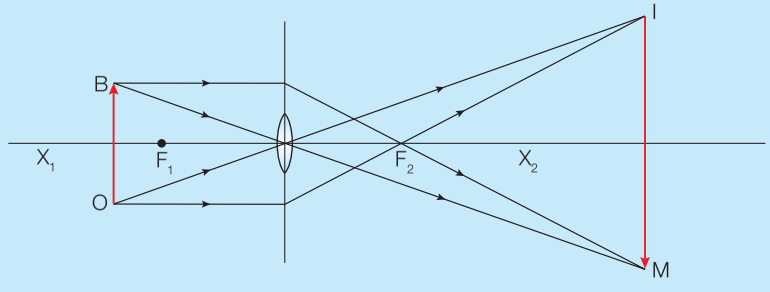
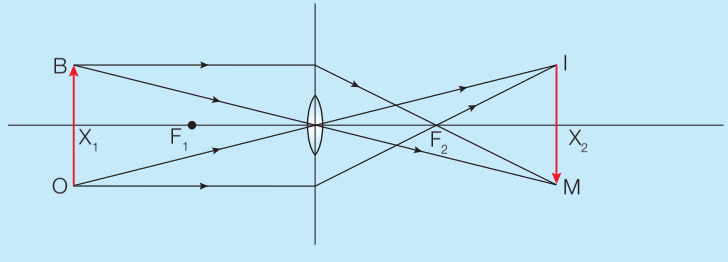
Note that the image is smaller than the object. We describe it as diminished or reduced. The image is also upright (again no surprise). If the same object is moved away from the lens, the image obtained will remain virtual, since the emergent rays will still be diverging, but will give an image further away from the lens than before and smaller. Note, also, that the image is never further away from the lens that the principal focus on the side of the lens from which the rays are incident.

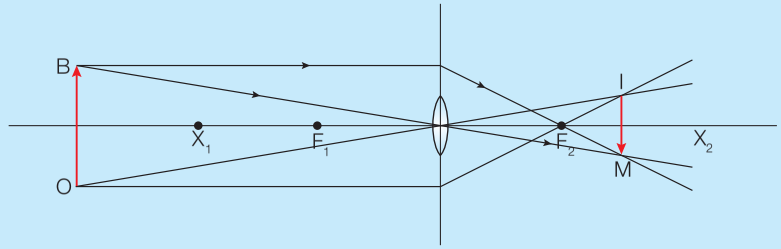
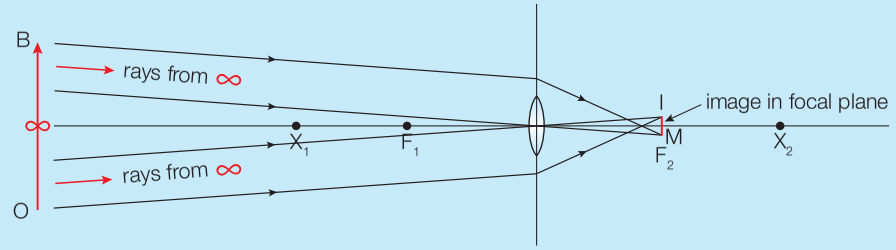
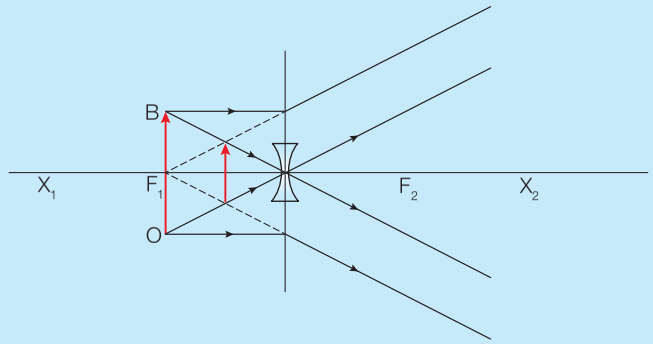
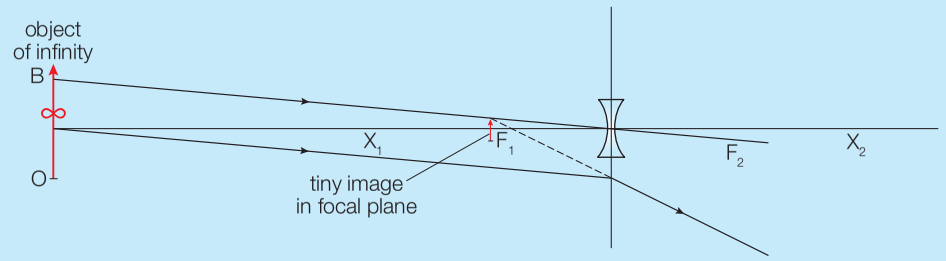
The differences between images produced by converging and diverging lenses

Note the differences between the features of the image produced by a converging lens and those of the images produced by diverging lenses. Whereas the diverging lens will always produce virtual images (which are upright and diminished, regardless of the position of the object, the converging lens will produce either real images (which are always inverted but could be either magnified or diminished) or virtual images (which are always upright and always magnified), depending on where the object is placed relative to the lens. Note, importantly also, that as long as the image is virtual it will be the same way up as the object (or upright), and on the same side of the lens as the object, whether the lens is converging or diverging. Table 23.3 gives a summary of the features of the images produced by the two types of lenses, converging and diverging.

A lens is any device that will cause parallel rays passing through it either to converge or to diverge.

Table 23.3 Features of images produced by converging and diverging lenses.

Type of lens	Object position	Image position	Image magnified or diminished?	Image real or virtual?	Image upright or inverted?	
converging	1 Between F_1 and C (where C is the optical centre)	on the same side of the lens as the object	magnified	virtual	upright	
	2 At F_1	at infinity	highly magnified	either virtual or real	either upright or inverted	
	3 Between F_1 and X_1, where $CX_1 = 2f$	on the other side of the lens between X_2 and infinity, where $F_2X_2 = f$	magnified	real	inverted	
	4 At X_1	on the other side of the lens at X_2, where $CX_2 = 2f$	same size as the object	real	inverted	

Type of lens	Object position	Image position	Image magnified or diminished?	Image real or virtual?	Image upright or inverted?	
	5 Between X_1 and infinity	on the other side of the lens between F_2 and X_2	diminished	real	inverted	
	6 At infinity on the side of F1	at F_2 in the focal plane	much diminished (tiny)	real	inverted	
diverging	7 Anywhere between C and infinity	on the same side of the lens as is the object between the principal focus and C	diminished	virtual	upright	
	8 At infinity	in the focal plane of the lens on the same side of the lens	very much diminished	virtual	upright	

Magnification of an image produced by a converging or a diverging lens

magnification ›

How much bigger or smaller an image is than the object is measured by the **magnification** of the image. The magnification is defined as

$$\text{magnification, } m = \frac{\text{height of the image}}{\text{height of the object}} = \frac{h_{im}}{h_{ob}}$$

Looking again at figure 23.18 we can see that, since the shaded triangles are similar, it follows that

$\frac{h_{im}}{h_{ob}} = \frac{v}{u}$, where v = image distance and u = object distance

In calculations, it is often more convenient to use the distances u and v to find magnification rather than image height and object height, since these distances are already provided in the data. One does not need to know the height of the object or that of the image to calculate the magnification. This formula holds good for images formed by both converging and diverging lenses.

The magnification of an image, $m = \frac{\text{image distance, } v}{\text{object distance, } u}$

Solving lens problems using a scale construction

Here is an example of a problem that can be solved by a scale construction:

A converging lens of focal length 20 cm forms an image of an object placed 30 cm from the lens. Determine by construction:

(i) the position of the image;

(ii) the magnification of the image.

To solve this type of problem, we represent the object and image distances by lines drawn to scale and we use the rules for constructing ray diagrams to find the image. We also adopt a scale for the object and image heights, especially since we shall be concerned with linear magnification.

The scale chosen for the object and image distances (u and v, respectively) depends on the width of the paper you will use. In choosing a scale for these distances, avoid using factors such as 3, 6, 7, 9, 11 and so on, which present difficulty when we have to divide by them.

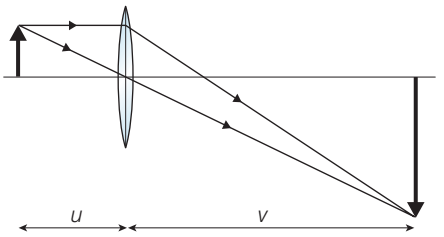


Figure 23.21

1 Draw a rough ray diagram (as in figure 23.21) to get some idea of the distances involved. The distance v looks about 3–4 times larger than distance u, so if u is 30 cm, then v is about 120 cm (a deliberate overestimate). This means that the distance between object and image is roughly 150 cm.

2 Choose a scale. If the paper has, for example a length of 30 cm along the long side, and we want to use as much of this length as we can (for a large diagram), we might consider using 1 cm of paper to represent 5 cm of object distance.

3 Choose a scale for the height of the object and image. This need not necessarily be the same scale as we used for the 'horizontal' part of the drawing. Remembering that the magnification is about 3–4, and that the image will be on the opposite side of the axis to the object, we are going to need space for about 4–5 units of length. Assuming that we have 20 cm to accommodate 5 units, we could take 1 unit of height to be 4 cm.

Construct the diagram as follows:

1 Draw up the axes as shown in figure 23.22. Draw the lens at the intersection of the axes.

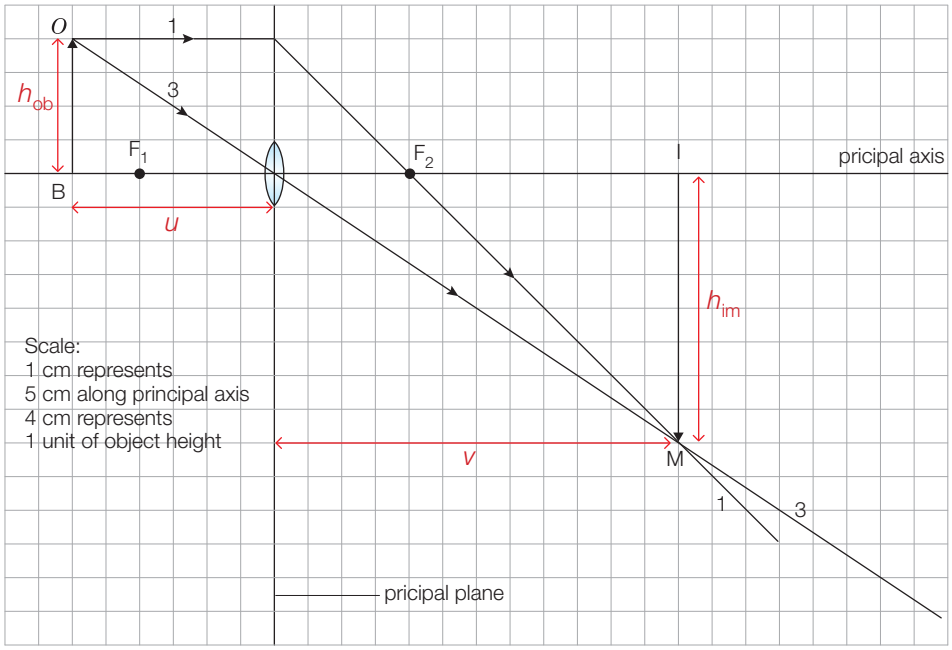


Figure 23.22 Your scale construction of a ray diagram should look something like this.

The vertical scale for the heights does not need to be the same as the horizontal scale for the distances.

1 Draw the upright object arrow on the principal axis 6.0 cm from the lens and 4.0 cm high. Put in the principal focus 4.0 cm from the lens on both sides.

2 Draw rays 1 and 3. The point of intersection of these rays is the image, M, of the tip of the arrow.

3 Draw the perpendicular from M to the principal axis of the lens. This perpendicular represents the image of the arrow.

4 Measure the length represented by v in the diagram, in centimetres.

5 Measure the height, h_{im}, of the image, in centimetres.

It can be seen from the construction that the image is magnified, inverted and real.

As the distance scale of the construction is 1:5, the measured distance v is multiplied by 5 to find the position of the image, i.e. the image is $5v$ cm from the lens.

Image distance = 5 × 12 cm = 60 cm.

The height of the object was drawn as 4 cm. The magnification is therefore $h_{im}/4$. (Magnification has no units, since we are dividing similar quantities.)

$$\text{Magnification} = \frac{8\text{ cm}}{4\text{ cm}} = 2$$

Points to note are as follows:

Reminder: as long as an image is real, it is inverted relative to the object; as long as the image is virtual it is upright relative to the object.

- The length needed across the page to accommodate $(u + v)$ was deliberately overestimated. This was to make sure that the 'image' would be formed on the graph sheet.
- A generous scale was chosen along the height axis. This was done to reduce the likelihood of the rays 1 and 3 being nearly parallel when they left the lens. It is difficult to tell quite where two lines meeting at a small angle intersect. The use of large objects and images reduces the uncertainty.

We take one more example of the use of a construction method to solve a lens problem

Worked example 23.5

A certain lens forms a real image of an object magnified four times when the object is placed 40 cm from the lens. Find by scale drawing the focal length of the lens see figure 23.23).

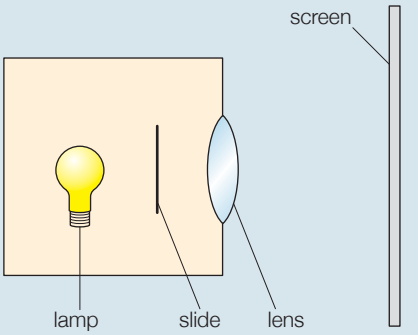


Figure 23.23

Solution

First you need to draw a rough diagram to guide you. This diagram is shown in figure 23.24. Since the image is real, the lens must be converging. Diverging lenses do not form real images.

So begin by setting up the principal axis and the principal plane using the information provided by the guide diagram that $u + v = (40 + 160)$ cm = 200 cm.

Assuming that your page is about 20 cm × 28 cm, use the long side for the x-axis.

Adopt a scale of 1 cm of paper = 10 cm of actual distance along the x-axis. Your scale up the page should be large for the sake of accuracy. So take 2 cm of paper to represent 1 unit of actual height of object or image.

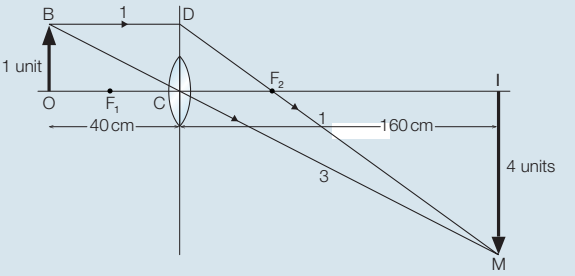


Figure 23.24 In solving a lens problem by construction, first draw a rough diagram showing how the image is formed.

The finished diagram shown in figure 23.24 was obtained as follows. Having set up the principal axis, the principal plane and the optical centre, the object OB was positioned 4 cm (40 cm) to the left of the optical centre.

The steps taken used to obtain the answers were:

1 Draw a line PQ 2 cm above the axis and another line RS 8 cm below the axis (Why? Because $m = h_{im}/h_{ob}$ and real images are always inverted about the axis.)

2 From B, draw ray 3 through the optical centre, C. This chief ray met the line RS at M, the tip of the image.

Now that we have located the tip of the image, we can now draw in ray 1.

1 Draw ray 1 from B to meet the principal plane at D and then join D to M. So ray 1 is BDM. But ray 1 passes through the principal focus of the lens. So the point at which DM crosses the principal axis is the principal focus of the lens. The value of the focal length of the lens from the construction is 32 cm.

You may wish to check this result by using the lens formula which we now introduce.

The lens equation (lens formula)

We have seen that problems involving the positions of objects and images may be solved by using scale diagrams. The same problems can be solved, perhaps more quickly, by using the lens equation or the lens formula, as it is sometimes called. The lens equation says that if a converging or a diverging lens forms an image of an object at a distance v from the lens, the object distance being u, then:

$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$, where f represents the focal length of the lens.

The full statement of this formula is the 'Real-is-Positive' or 'R-P' lens formula.

In using the formula, we must know the signs (positive or negative) to be given to each of the quantities u, v and f. We give a positive sign to distances associated with real objects, real images and real principal foci and a negative sign to distances relating to virtual objects, virtual images and virtual principal foci. We will never meet virtual objects at this level, and so u, the object distance will always be positive. In effect, we associate 'realness' with a positive sign and 'virtualness' with negative sign. Here are three examples.

Worked example 23.6

A lens placed 15 cm from an object produces a real image 40 cm from itself. Calculate the focal length of the lens and state whether the lens is converging or diverging.

Solution

Points to consider:

- Since the object is real, the object distance, u, is +ve.
- Since the image is real, the image distance, v, is +ve.
- We do not assume a sign for the focal length of the lens. This will be revealed in the answer. We expect the answer to be +ve, since real images of objects are produced only by converging lenses. A positive result for the focal length will show that the lens used is converging.

So applying the 'RP formula' we have

$$\frac{1}{+15} + \frac{1}{+30} = \frac{1}{f}$$

That is, $\frac{1}{f} = \frac{3}{30} = \frac{1}{10}$

Giving $f = +10$

So the value of the focal length is +10 cm and since its sign is +ve, the focal length is +ve, the principal focus is real and the lens in question is converging.

Worked example 23.7

A lens forms a virtual image of a real object placed 40 cm from itself. The distance of the image from the lens is 60 cm. Determine the nature and focal length of the lens.

Solution

Since the image is given as virtual, its distance from the lens, v, is negative; the object is real and so its distance from the lens is +ve.

So here $u = +40$ cm; $v = -60$ cm and so we have $\frac{1}{+40} + \frac{1}{-60} = \frac{1}{f}$

that is, $\frac{1}{f} = \frac{3}{120} - \frac{2}{120}$

or $\frac{1}{f} = \frac{1}{120}$

giving $f = +120$

So, since f is positive, the lens is a converging lens and its focal length is 120 cm.

Worked example 23.8

How far must an object be placed from a diverging lens of focal length 20 cm if it is to form an image of the object with a magnification of ½?

Solution

Here we must remember that a diverging lens will always produce virtual images of real objects. If the image here is virtual, then the image distance is –ve and if the object is real, its distance from the lens is +ve. The magnification may be taken as v/u and if v is –ve, the ratio v/u will be –ve.

This means then that $v/u = -1/2$ and that $v = (-1/2)u$ and so $\frac{1}{v} = \frac{-2}{u}$.

Substituting in the lens equation, $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$

we have $\frac{1}{u} + \frac{-2}{u} = \frac{1}{f}$

that is $\frac{-1}{u} = \frac{1}{f}$

since the focal length = 20 cm, we have

$\frac{-1}{u} = \frac{1}{-20}$ (note that f is –ve, since the lens is diverging)

and so $u = 20$ and $v = (-½)u = -10$ cm.

The object distance = 20 cm and the object must be placed 20 cm from the lens.

Check: magnification, $m = v/u = ½$, then since $v = ½u$, v should therefore be 10 cm, which is correct.

Worked example 23.9

Where should an object be placed on the axis of a lens, in order that the image formed by the lens might be as large as the object? The focal length of the lens is 30 cm.

ITQ5

How would you find that mystery point on the graph of *v* against *u* where *v* = *u*?

ITQ6

Assuming that you have found the point on the graph where *v* = *u*, use it to find the focal length of the lens used in the activity

Solution

We must remember that all images formed by diverging lenses are diminished. The lens in question must, therefore, be converging.

If the object and image are of the same size, then $m = \frac{v}{u} = 1$ and so $v = u$.

If we need to find u, then u must be preserved in the lens equation and, substituting, we have

$$\frac{1}{u} + \frac{1}{u} = \frac{1}{f}$$

or $$\frac{2}{u} = \frac{1}{f}$$

giving $$u = 2f$$

$$= 2 \times 30\,\text{cm}$$

$$= 60\,\text{cm}$$

MATHEMATICS: simultaneous equations

To close the chapter, let's go back to part (ii) of Practical activity 23.1, where we plotted a graph of u against v. You most likely got a curve when you did. If you can find the point on that curve where $v = u$, you can find the focal length of the lens which gave that curve.

If we take the other graph you plotted in Practical activity 23.1, of $1/v$ against $1/u$, you would most likely have obtained a straight line. We know now that $1/u + 1/v = 1/f$. The equation of the graph you plotted was $1/v = -(1/u) + 1/f$. What are the answers to ITQ7?

ITQ7

(a) What is the intercept on the $(1/v)$-axis?

(b) What is the intercept on the $(1/u)$-axis?

(c) Should these intercepts be equal? Were they in your graph?

(d) What was the value of each intercept for your graph?

(e) How can the focal length of the lens used in the activity be found from the graph?

Chapter summary

- A lens is any device that will cause parallel rays passing through it either to converge or diverge.
- Lenses may be considered to be made up of truncated prisms of continuously increasing (or decreasing) refracting angle, or vice versa.
- Lenses may be classified as either converging or diverging, depending on their effect on parallel light rays.
- Important ideas used in ray diagrams for images formed by lenses are:
  - principal axis;
  - optical centre;
  - principal plane;
  - principal focus;
  - focal plane;
  - paraxial rays.
- A converging lens has two real foci and a diverging lens has two virtual foci, one on each side of the lens.
- A converging lens can produce both real and virtual images of an object: real if the object is placed beyond the principal focus, but virtual if the object is placed between the principal focus and the lens.
- Real images produced by converging lenses may be either diminished or magnified; virtual images produced by converging lenses are always magnified.
- A diverging lens always produces a virtual, diminished, upright image of an object regardless of the position of the object.
- The real images produced by converging lenses are inverted about two perpendicular planes. Such images are said to be 'twice inverted'.

Answers to ITQs

ITQ1 Having drawn the mirror image in the usual way (as for a plane mirror), if the diagram is rotated through 180° about the principal axis, you will obtain the same diagram in all planes drawn through that axis

ITQ2 If you use the chief ray through the point O it will coincide with the principal axis and so the other ray used to define the position of the image of O must meet this chief ray on the principal axis. Hence the image of O must be on the principal axis.

ITQ3 If a ray diagram is drawn for an object lying in a horizontal plane it must look the same as one for a vertical plane, since the image in all planes through the principal axis will be mirror image about the principal axis. Instead of using an object standing on the principal axis, one could just as well use one lying with O touching the axis. The image would be the mirror image about that axis.

ITQ4 Consult diagram 5 of table 23.3 for the answer.

ITQ5 On the same axes as were used for plotting the graph of v against u plot the graph of $v = u$. The 'mystery point' is where the two graphs intersect. Read off the value of u or v. This value will be the solution.

ITQ6 Use the lens equation. When $u = v$, each of these $= 2f$. So $f = v/2$ or $u/2$

ITQ7 Intercept $= 1/f$ in each of **(a)** and **(b)**, since the equation of the graph is $1/u + 1/v = 1/f$ and, rearranging, $(1/v) = (-1/u) + (1/f)$.

When $\frac{1}{v} = 0$, intercept $1/u = 1/f$; and when $1/u = 0$, intercept $1/u = 1/f$.

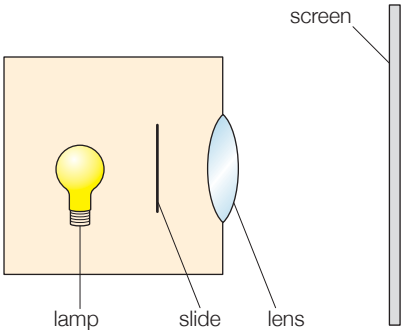
(c) Yes, the intercepts should be equal.

(e) Take the mean of the two intercepts. The focal length will be the reciprocal of this mean.

Examination-style questions

1 (i) With the help of a labelled diagram, explain what is meant by the focal plane of a lens.

(ii) The diagram represents a slide projector that is being used for producing a magnified image of a slide on a distant screen.



(a) Explain the function of each of the labelled parts in the diagram.

(b) Explain why the distance between the lens and the slide must always be slightly, but not too much, greater than the focal length of the lens. What would be the effect on the image if this distance were made:

1 less than the focal length of the lens?

2 much greater than the focal length of the lens?

(iii) What adjustment must be made to produce an image on a more distant screen? What effect will this have on the size and brightness of the image on the screen?

(iv) When a lens is used to produce an image on a screen, the image is always twice inverted. What is meant by *twice inverted*?

2 A certain converging lens produces a real image of an object with a magnification of 4 when the object distance is 20 cm. How far from the lens should the same object be placed to produce a virtual image with the same magnification of 4?

Section D:

Electricity and Magnetism

24 Electrostatics

By the end of this chapter you should be able to:

- describe experiments to demonstrate charging by friction
- describe simple experiments to show that like charges repel and unlike charges attract
- explain the charging of objects in terms of properties of negatively charged electrons which are relatively free to move
- distinguish between conductors and insulators
- explain how a charged object can attract objects having zero net charge
- describe and explain charging by friction, contact and induction
- describe how a gold-leaf electroscope can be used to detect and estimate charge
- describe the distribution of charge on a conductor of variable shape
- define an electric field as a region in which a charge experiences an electric force
- represent diagrammatically the electric fields around and between point charges and between charged parallel plates
- describe the structure of a simple capacitor
- define the farad, the S.I. unit of capacitance, in terms of charge
- describe hazards and useful applications of static electricity



Static and current electricity

current electricity ›
charged particle ›

The electricity that comes into our homes along wires is called **current electricity**, and it consists of a continuous flow of 'charged' particles. **Charged particles** may flow in solids (as along wires), in liquids (as in car batteries), in gases (as in fluorescent lamps) or even in a vacuum (as in TV picture tubes).

static electricity ›

Electricity consisting of 'charged' particles that are at rest is called **static electricity**. Static electricity may give rise to current electricity, as happens during a lightning storm (figure 24.1). However, such currents are of very brief duration.

The word electricity comes for the Greek word *elektron*, which means 'amber'. Amber is a hard, yellow, fossilised tree resin.



Figure 24.1 Lightning is caused by static electricity.

Static electric charge

When a plastic pen barrel is rubbed against cloth, it can attract small pieces of paper (figure 24.2). We say that the plastic has become charged. The ancients knew of this phenomenon. They knew that when you rub amber with cloth, the amber attracts dust. In other words, the amber becomes 'charged', or as we would say today, it has acquired 'static'.

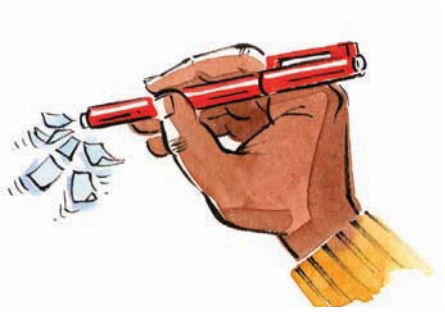


Figure 24.2 A plastic pen that has been rubbed with cloth becomes charged and attracts pieces of paper.

Practical activity 24.1

Demonstrating static electric charge

1 Rub a plastic pen vigorously on your clothing.
2 Quickly place the rubbed pen close to a tiny piece of paper (or aluminium foil). The paper (or foil) is attracted to the pen (because the pen has become charged) and then jumps away.
3 Now a fun part. After rubbing the pen vigorously on your clothing, place it close to (but not touching) a dry, empty soda can lying on its side on a level, smooth table. The can will roll toward the pen, because the pen has been charged and is attracting the can!

Law of charges

Are charges all of the same kind? Figure 24.3 shows that a plastic pen rubbed with cloth repels a similarly rubbed plastic pen that is suspended with a thread. A glass rod rubbed with silk cloth repels another glass rod similarly rubbed. However, a glass rod rubbed with silk attracts a plastic pen rubbed with cloth.

The above experiments, and others, show that charges of a similar kind repel each other. Charges that are unlike attract each other. This is a fundamental law of electrostatics.

Like charges repel; unlike charges attract.

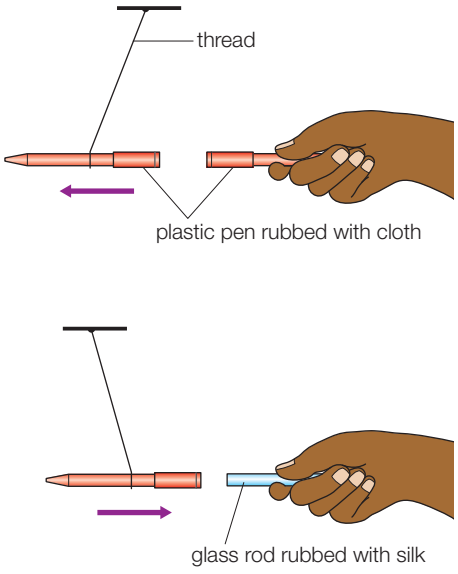


Figure 24.3 Repulsion and attraction of like and unlike charges.

Practical activity 24.2

Investigating electrostatic attraction and repulsion

(Hint: an arrangement such as shown in figure 24.3 can be used.)

In a group do the following:

1 Design and carefully carry out activities to find out if:

(a) a charged plastic pen repels a similarly charged plastic pen;

(b) a charged plastic pen attracts or repels a similar plastic pen that is uncharged;

(c) an uncharged plastic pen attracts or repels a similarly uncharged plastic pen.

2 Discuss whether your observations agree with the fundamental law of electrostatics.

Further experiments have shown that there are only two kinds of charge. The type found on the rubbed pen has been called 'negative charge'. The type found on the rubbed glass rod has been called 'positive charge'. Nowadays, we can define a **negative charge** as the net charge on a polythene strip that has been rubbed with a woollen cloth. We can define **positive charge** as the net charge produced on a Perspex rod that has been rubbed with a silk cloth.

negative charge ›
positive charge ›

ITQ1

Will a polythene strip rubbed with a woollen cloth attract or repel a plastic pen that has been rubbed with cloth?

Electrostatic photocopier

The attraction between positive and negative charge is used in the electrostatic photocopier, which is now described:

- A rotating drum is connected electrically to the Earth. The surface of the rotating drum, covered with a special 'photoconducting' material (figure 24.4), is first given a positive charge.
- A strong light is moved across the paper that is being copied. The dark areas on the paper absorb the light, but the white areas reflect the light on to the drum as it rotates.
- Where light falls on the drum, the surface becomes 'conducting' and the positive charge in these areas are discharged (neutralized) by negative charges flowing from the Earth (see section titled 'Discharging by earthing' on page 376. The pattern of positive charge on the drum then corresponds to the pattern of dark areas on the paper being copied.

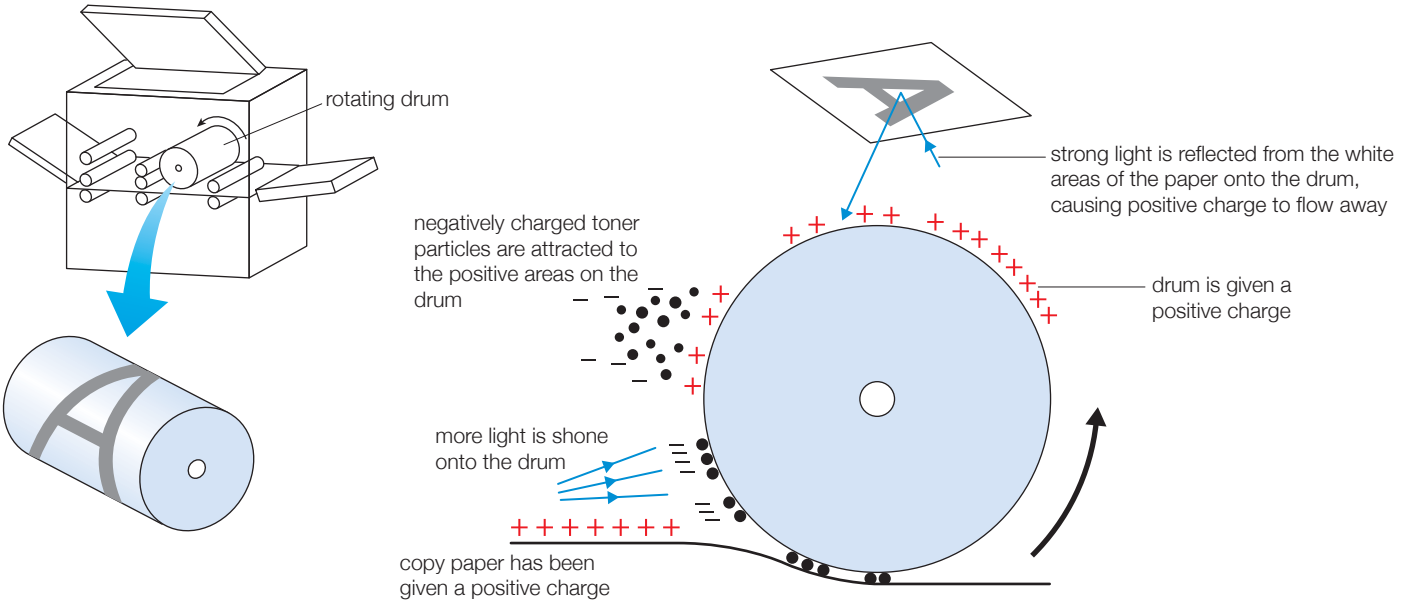


Figure 24.4 The electrostatic photocopier.

- Dry toner (ink) particles from the toner cartridge are given a negative charge, and the negatively charged toner particles are attracted to the pattern of positive charge on the drum as it moves round.
- More light is shone on to the drum immediately after the toner pattern has been created on it. This causes further loss of any excess positive charge remaining on the drum.
- However, the copy paper is given a strong positive charge before it comes into contact with the drum. The positively charged copy paper attracts the negatively charged toner particles. The toner pattern on the drum is pressed and baked on to the copy paper, reproducing the original pattern.

Electrostatic spray painting

Internet search term: electrostatic painting

In electrostatic spray painting, the metal body of a car is connected electrically to a negatively charged source (or to the Earth). The spray gun gives off fine paint droplets which are positively charged. The positively charged paint drops are attached very strongly to the metal body due to the attraction of unlike charges. A smooth finish is obtained since the droplets, having the same kind of charge, repel each other and not clump together.

Forces between charged particles

coulomb ›

The S.I. unit of electric charge is the **coulomb** (C), named after the French scientist, Charles Coulomb, who investigated the forces between charged particles.

Coulomb found that the force, F, between two tiny charges, Q_1 and Q_2, was directly proportional to the product of the two charges. Thus, for a given distance, r, between them, if Q_1 was doubled, the force was doubled. If Q_1 was doubled and Q_2 was trebled, the force became six times the original force.

Coulomb also found that the force between two charges separated by a distance, r, was proportional to $1/r^2$ (in other words, the force varies inversely as the square of the separation distance). Thus, if r was trebled, the force *reduced* to $1/3^2$ or 1/9 of its original value. If r was halved, the force *increased* to $1/0.5^2$ or four times its original value.

ITQ2

A force of 40 N exists between two charges separated by a distance of 2 m. What will be the new force if the charges are separated by a distance of 8 m?

Electron theory of charging and conduction

According to the atomic theory, all substances are made up of tiny particles called atoms. (The atomic theory of matter is dealt with in more detail in chapters 32 and 33.) The central part of the atom, called the **nucleus**, contains subatomic particles with a positive charge, called **protons**, and uncharged particles called neutrons. Although **neutrons** have no charge, their presence helps to keep the positive charges together in the nucleus. Orbiting the nucleus are very light particles, negatively charged, called **electrons**.

nucleus ›

proton ›

neutrons ›

electron ›

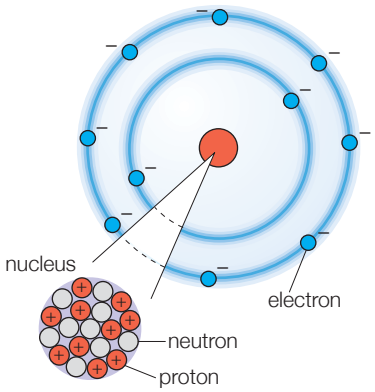


Figure 24.5 Arrangement of charged particles in an atom.

Substances do not usually appear to have a charge because each of their atoms has equal numbers of positive charges (protons) and negative charges (electrons). These together make each atom neutral (figure 24.5).

ITQ3

Explain, in terms of movement of charge, what kind of excess charge a woollen cloth acquires after being used to rub a polythene rod.

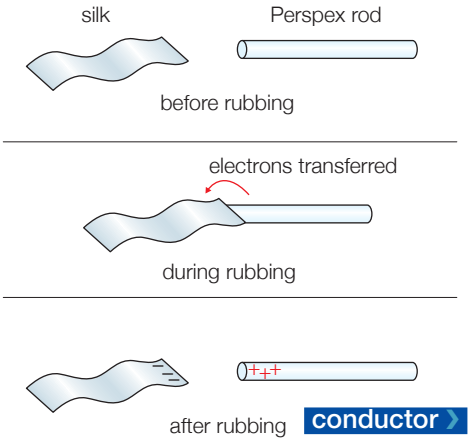


Figure 24.6 Charging involves movement of electrons.

CHEMISTRY: electrons; properties of metals and non-metals

Pages 507–511 contain an explanation, in terms of filled orbits, as to why electrons in outermost orbits may be relatively 'free' or 'tightly bound'.

Electron theory and charging

When electrons are removed from a neutral atom, the atom is left with more positive than negative charges. The material made up of these atoms thus becomes positively charged. If electrons are added to the atoms of a material, there are more negative than positive charges and the material is negatively charged.

It is easier to move electrons from an orbit than positive charges from a nucleus. Thus, charging usually involves movement of electrons, whether the substance acquires a net positive or net negative charge (figure 24.6).

Electron theory and conduction

conductor

In the atom of some substances, called '**conductors**' (figure 24.7 (a)), the outermost electrons are few and are held loosely to the rest of the atom. These electrons can move freely from atom to atom and are called 'free' electrons. When a positively charged rod is brought near to one end of an object made of a conducting material, 'free' electrons are attracted and rush towards the part of the object near the positive charge end of the rod. Such materials therefore conduct electricity very well. Silver and copper are among the best conductors. All metals are good conductors of electricity since they all have 'free' electrons.

insulator

In the atoms of materials known as '**insulators**' (figure 24.7 (b)), the electrons are held tightly in their orbits. These electrons cannot move freely from atom to atom. Hence, insulators do not normally conduct electricity. When a positively charged rod is brought near to one end of an insulator, such as a piece of paper, the electrons remain with their parent atoms. However, the orbits of the electrons might be *displaced* towards the positively charged rod (see figure 24.8).

However, if the charge next to the insulator is very great, then electrons may be torn from their parent atoms, and conduction may take place within an insulator. When this happens, we say that the insulation has 'broken down'.

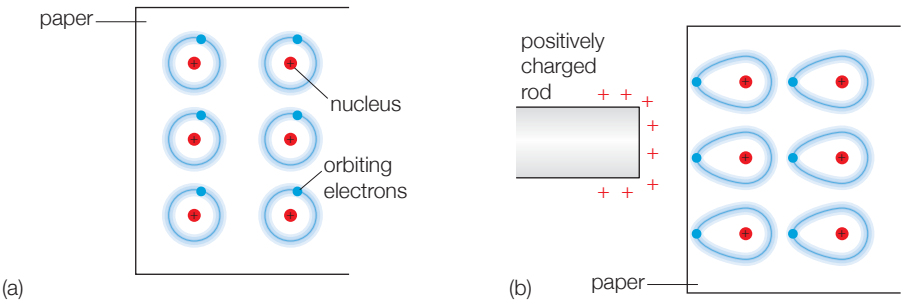


Figure 24.8 (a) Circular orbits of electrons in a piece of paper (atoms are drawn simplified). (b) Orbits being displaced by the presence of a positively charged, nearby rod. The part of the paper near to the rod therefore becomes somewhat negatively charged and can be attracted by the object.

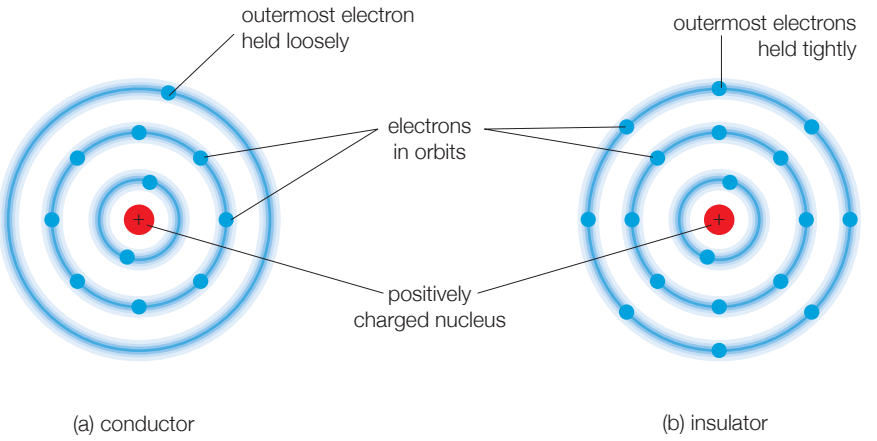


Figure 24.7 Models of an atom of a conductor and of an insulator. (a) Conductor – outermost orbit not full to capacity, (b) insulator – outermost orbit full to capacity.

Sparks are generally produced when insulation breakdown occurs, as in a car spark plug (figure 24.9).

The electron theory also helps to explain why the charged pen shown in figure 24.2 attracts bits of paper that are not charged. The part of the paper nearest the positively charged pen becomes slightly negative, as shown in figure 24.8, and is therefore attracted to the pen. Thus, a charged object can attract one that has zero net charge! In other words, one cannot conclude that if attraction takes place, the two objects must have had unlike charges to start with. The only sure test for the sign of the net charge on an object is therefore the repulsion test. If an object repels a second object, then the net charges on both must have been like.

When the paper attaches to the pen, a few electrons might be pulled from atoms of the paper and join atoms of the pen. The paper would then have excess positive charge and be repelled from the positively charged pen. (Of course, the pen could also lose its excess charge in a humid atmosphere. When this happens, the paper simply falls off the pen.)

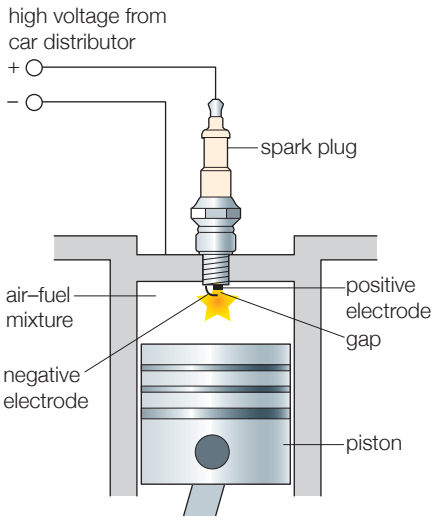


Figure 24.9 The spark plug inside a motor car engine cylinder sparks when the insulation of the gases present 'breaks down'. The breakdown occurs in a small gap between two electrodes when one electrode is made very positive compared with the other. Electrons are pulled off atoms of the air–fuel mixture and move rapidly to the positive electrode. A large current thus flows during the breakdown, producing sparks. The sparks ignite the fuel-air mixture in the cylinder, which provides power to keep the engine running.

Methods of charging and discharging

Materials become charged, or in other words have excess positive or negative charge, by removal or addition of electrons. A charged substance can also become discharged by movement of electrons.

Charging by contact

A negatively charged, insulated metal rod is brought near to a *light* metal ball that is suspended by an insulating thread (figure 24.10). Electrons are repelled to the far side of the metal ball. (Remember, electrons can move freely in metals.) The side of the ball nearest the metal rod is left with an excess of positive charge and is attracted to the negatively charged rod.

When the ball and the rod touch, electrons from the rod are attracted and rush towards the positive charges on the ball. The ball now has excess negative charge and is repelled by the negatively charged rod. The negative charge on the rod has been shared between the ball and the rod.

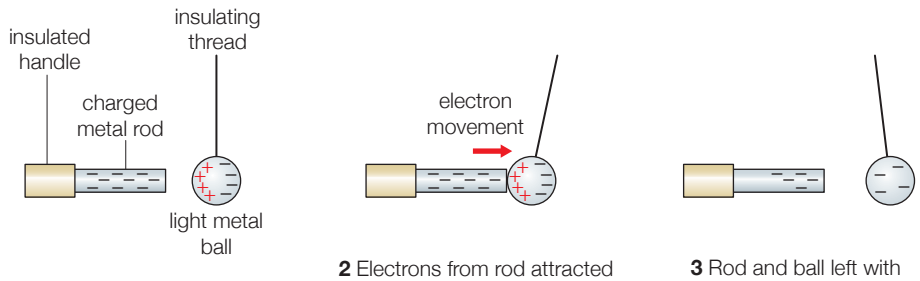


Figure 24.10 Charging by contact.

This method leaves the charged object with the *same kind of charge* as that on the charging object. Charging by contact works very well with conductors: charge is easily shared between the conductors because they have free electrons that can move readily.

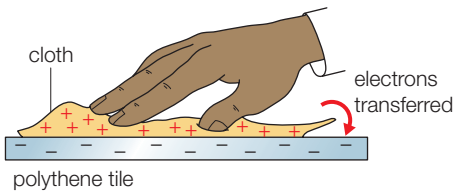


Figure 24.11 Charging by friction: rubbing a polythene tile with a woollen cloth.

Charging by friction (rubbing)

When a polythene tile (figure 24.11) is rubbed vigorously with a woollen cloth, a few electrons are removed from the cloth to the tile. The tile thus becomes

negatively charged (excess electrons). The cloth becomes positively charged (having given up some electrons).

Charging by friction works best with insulators or insulated conductors. Such charged objects tend to remain charged since there are no 'free' electrons to carry charge to or from them. Friction between two such objects results in their becoming oppositely charged.

Charging by friction has useful applications. When the flints in flint lighters are rubbed together, sparks are produced because the flints become so highly charged that electrons are ripped off atoms in the air between them. The insulations of the air breaks down and an electric current flows briefly. The air heats up rapidly and sparks are produced. The sparks are used to ignite the gas from Bunsen or stove burners.

Did you know?

Charged objects readily attract bacteria. Floors in operating theatres are therefore made of conducting material and earthed. This prevents build-up of charges as shoes or carts rub against the floor.

Discharging by earthing

discharged ›
earthed ›

Charging by friction can also create dangers. Trucks transporting fuels can become charged and generate sparks as they rub against the air when moving. Sparks in the vicinity of fuel can be dangerous. These vehicles are therefore equipped with conducting tyres so that the charges acquired can **discharge** harmlessly to the ground. This is called '**earthing**' the vehicle. Some fuel trucks have a connection such as a chain that drags along the ground for this purpose.

Internet search term: aircraft refuelling
Internet search term: static electricity fires at gas stations

Aircraft can likewise become charged as they move through the air and on the ground, especially during landing. Sparks, jumping from a highly charged aircraft on to metal objects in the vicinity of aviation fuel vapour, can ignite the vapour. As a safety precaution, therefore, aircraft are connected to the Earth to ensure that they are discharged before refuelling commences.



Figure 24.12 Gas station fire.

Gas station fires
It has been reported in the USA that many self-service gas station fires have started when a motorist, upon sliding out of the car seat, went to take the gas pump out of the car after refuelling had just finished (figure 24.12). How could the fire have started? What could have been done to prevent it? How could occurrences like these be prevented from ever happening?

How does earthing help to discharge an object? The planet Earth can be regarded as an enormous conductor. When a charged metal object is connected to it, the charge is shared between the two by movement of electrons. The Earth, being so large, receives practically all of the charge, leaving the object with practically none. We say that connecting an object to earth results in the object being discharged. A conductor with a small charge can similarly be discharged by touching it with a part of your body. Practically all of the excess charge goes to your body.

Charging by induction

induction ›

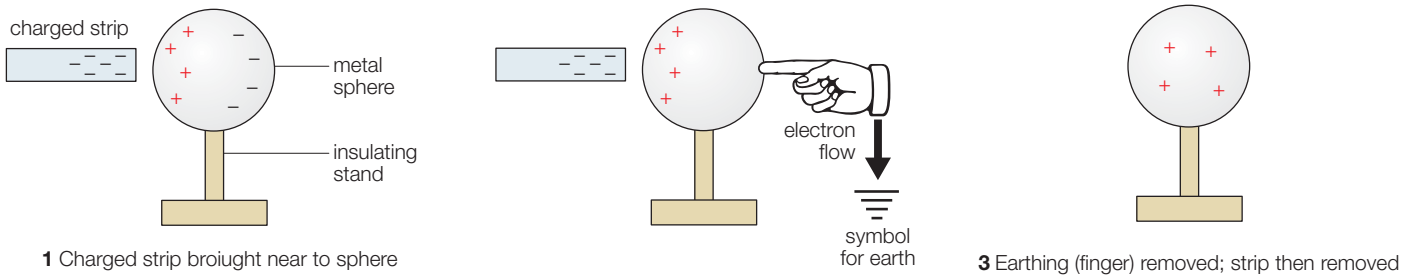
Figure 24.13 shows a negatively charged strip being brought near to a metal sphere. Electrons (negative charges) move to the far side of the sphere. This leaves the side next to the strip with excess positive charge. Since the strip and the sphere do not touch, we say that this separation of charges in the sphere has been induced, 'induced' here meaning 'brought about by an influence'.

If the sphere is now connected to earth by a wire, with the charged strip still near to the sphere, electrons move from the sphere to earth. When the wire is removed, the sphere is left with an excess of positive charge. A similar effect can be obtained by touching the sphere with the finger, as shown in figure 24.13, and then removing the finger with the strip remaining near to the sphere.

ITQ4

In figure 24.13, step 2 (charging by induction), electrons travel from the sphere to earth when the sphere is connected to earth. Why don't electrons travel back from the earth towards the excess positive charges that are left in the sphere?

When a single object is charged by induction, the result is an excess charge *opposite* to that on the charging object.



1 Charged strip broiught near to sphere

2 Sphere touched with finger or connected to earth by wire while strip remains near to sphere

3 Earthing (finger) removed; strip then removed

Figure 24.13 Charging by induction (earthing method).

Practical activity 24.3

Investigating charging by induction

In a group do the following:

1 Carry out the steps shown in figure 24.13, but using a positively charged strip.

2 Predict the sign of the charge that would be obtained in step 3, giving reasons.

3 Test the sign of the charge to see if your prediction is correct. (Hint: You may use a light object of known charge to test the sign of the charge.)

Detecting and estimating charge

A metal-leaf electroscope can be used to

- detect charge,
- distinguish between the two kinds of charge, and
- estimate the sizes of charges.

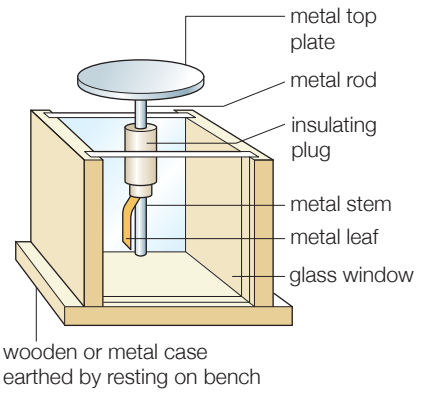


Figure 24.14 A metal-leaf electroscope.

The electroscope consists of a metal plate or cap connected to a metal rod (see figure 24.14). At the bottom of the rod and electrically connected to it, is a very thin metal leaf. The rod and cap are insulated from the case, which is usually connected to earth.

When a negatively charged rod, for example, is brought near to the metal cap of the electroscope, electrons from the cap move towards the leaf and the lower part of the rod. The leaf and lower rod become negatively charged, and the rod repels the leaf. The larger the charge brought near to the cap, the more the leaf diverges. The angle of divergence of the leaf can be calibrated (see page 15 for the meaning of 'calibrated') to indicate the size of charge detected. Thus, the electroscope can be used to show whether an object is charged and also whether one object has more charge than another.

IT Q5

An electroscope is charged positively (e.g. by induction). How can this charged electroscope be used to test the size of the charge on a rod?

Distribution of charge on the surface of a conductor

proof plane ›

The distribution of charge on the surface of a conductor can be investigated by touching the conductor with a small, insulated metal disc, called a **proof plane**, at various places on the conductor (figure 24.15). The disc receives a charge by contact that is proportional to the density of charge at the place where it touches the conductor. The charge on the disc is transferred to an electroscope through contact with the cap. The amount of charge that was taken up by the disc can be estimated from the size of the divergence shown by the metal leaf.

In practice, the disc is made to touch the inside of a hollow metal can placed on the cap of the electroscope. This ensures that all the charge is transferred.

It is found that the charge density is greatest at the most curved or pointed parts of the conductor.

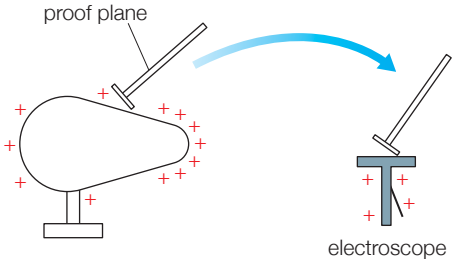


Figure 24.15 Investigating the distribution of charge on the surface of a conductor of variable shape.

Action of points

Figure 24.16 shows a positively charged, pointed conductor. Most of the charge is at the pointed end. Any negative ions present in the air will be easily attracted to the highly charged point. Electrons from these negative ions jump towards the highly charged point, causing the positive charges there to be neutralised. The point therefore quickly discharges. If negative ions are not present in large quantities, electrons may be pulled from air molecules that come in contact with the highly charged point. The electrons neutralise the positive charges, causing the point to discharge quickly. The positive ions that are produced in the air are repelled by the charged point until the conductor is discharged.

The electrostatic dust precipitator and the lightning conductor are two devices that make use of action of points.

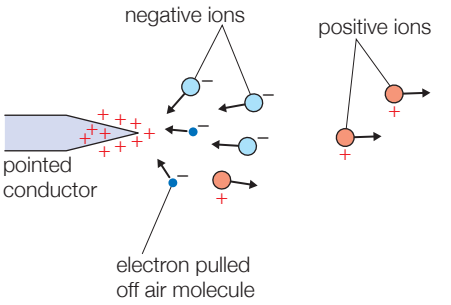


Figure 24.16 Action of a point, showing movement of positive ions, negative ions and electrons in the air.

ions ›

Ions are atoms (or groups or atoms) that have gained or lost electrons.

Dust precipitator

Figure 24.17 shows a dust precipitator inside a factory chimney. The anode (the positively charged component), made of fine mesh, charges the air in the chimney positively by action of points. On contact with the positive ions produced in the air, soot and dust particles become charged positively and are attracted to the cathode (the negatively charged component). As a result, the air becomes free of these solid particles. The soot and dust are periodically removed from the cathode by scrubbing.

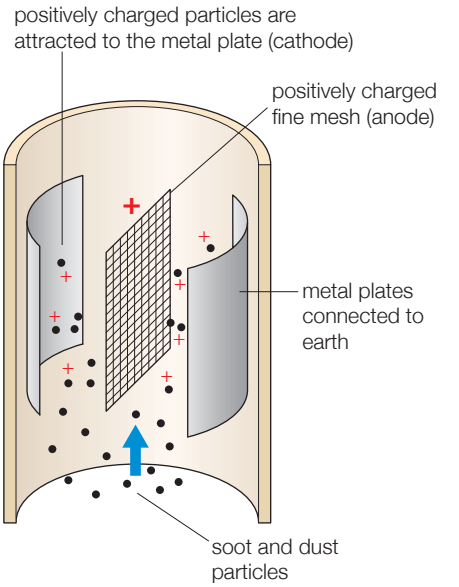


Figure 24.17 Electrostatic dust precipitator.

Internet search term: dust precipitator

Lightning conductor

Lightning is a gigantic spark occurring between a charged cloud and the Earth (figure 24.18), or between charged clouds. As electrons rush between them, they may knock more electrons out of atoms with which they collide. This provides an even greater rush of electrons. The collisions of electrons with atoms in the air produce heat and the brilliant flash of lightning. It is the sudden expansion of the heated air that causes thunder.

Lightning usually strikes the tallest and most pointed objects. For this reason, it is very unsafe to shelter under a tree during a thunderstorm. The points of these objects become highly charged by induction when a charged cloud passes overhead. Charge can then easily flow by action of points from the pointed object to the passing, charged cloud. If you were to touch the objects while charge was flowing you could be electrocuted.

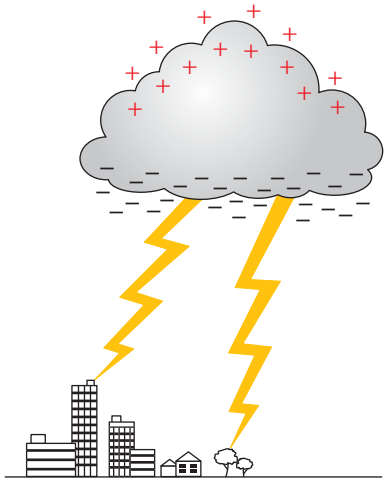


Figure 24.18 Did you know that one lightning flash can discharge as much as 3 000 000 kWh of energy – enough to power a small town for several days?

ITQ6

Parents usually advise their children not to hold sharp objects like knives or scissors during a lightning storm. Why is this good advice?

lightning conductor ›

A **lightning conductor** is a tall, pointed metal rod that is grounded (connected to earth). The top of the rod becomes charged through induction from a charged overhead cloud, and there is a high density of charge at the pointed ends (figure 24.19). Clouds discharge through such a conductor steadily, through action of points, in preference to other less pointed nearby objects. A massive build-up of charge at the top of the building, which could result in the insulation of the air being broken down, and a huge flash of lightning occurring, is prevented.

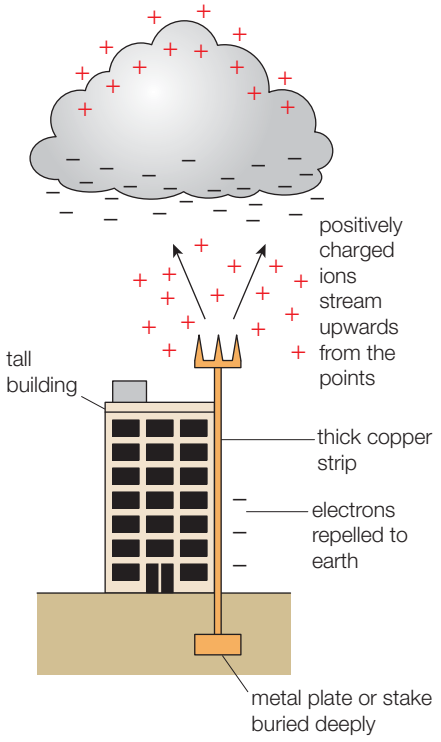


Figure 24.19 A lightning conductor provides a steady discharge through action of points.

Does lightning travel upward or downward?

Consider a thundercloud that has a high positive charge on top and a large negative charge at the bottom. A first flash can occur consisting of negative ions and electrons travelling from cloud to ground. The air could become very ionised during the first flash. This provides an abundance of ions for a second flash. This time, negative ions and electrons travel from ground to cloud to neutralise the positive charges on the cloud. In this flash, as much as one million amperes of current can flow!

Electric fields

electric field ›

We saw earlier in this chapter that electric charges exert forces on each other. we say that there is an **electric field** around each electric charge.

> **An electric field is the region in which an electric force acts.**

A charged object, placed in an electric field, experiences an electric force – the greater the strength of the field, the greater the force experienced.

Direction of an electric field

The direction of an electric field is defined as the direction in which a tiny positive charge would move if placed at any given point in the field. Field diagrams are shown below for various arrangements of positive and negative charges (figure 24.20).

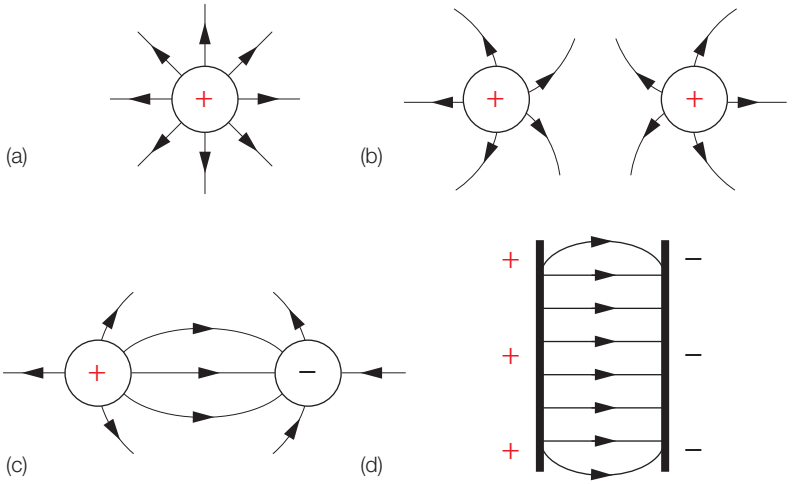


Figure 24.20 Electric field diagrams for various charge arrangements. The strongest fields are near point charges. (a) Positive point charge. (b) Two positive point charges. (c) A positive and a negative point charge. (d) Two parallel, conducting plates.

ITQ7

What would the electric field due to a negative point charge look like?

field lines ›

field strength ›

The **field lines** give an idea of the direction and relative strength of a field at various locations. The closer together the lines are, the stronger is the electric field in that region. For point charges, the **field strength** is greater near to the points. The field strength is constant throughout most of the space between two parallel, charged, conducting plates.

Storing electric charge

capacitor ›

Electric charge is stored in devices called **capacitors**. Electronic flashes in cameras make us of 'discharge currents' from charged capacitors. Electronic oscillators in radios and musical instruments use capacitors to store charge and discharge it rapidly. In the early days of electrostatic investigations, the **Van de Graaff machine** was used for storing large quantities of charge for experimental purposes.

Van de Graaff machine ›

Other kinds of fields:

- A gravitational field is a region in which a gravitational force acts (see chapter 6).
- A magnetic field is a region in which a magnetic force acts (see chapter 27).
- Field lines, similar to those in electricity, give the direction and relative strengths of these fields at various locations.

Capacitors

A capacitor is an arrangement of two conductors, A and B, separated from each other (see figure 24.22). The conductors may be separated by an insulator such as air, paper, plastic, ceramic, metal oxide, or even by a vacuum.

A capacitor is charged by moving electrons from one of the conductors to the other. A capacitor can easily be charged by

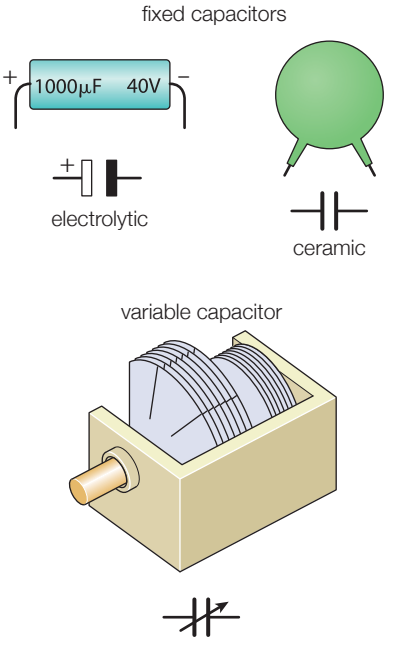


Figure 24.21 Capacitors can be found in electronic equipment such as radios, computers and cameras. The corresponding electrical symbols are shown for the different types of capacitor.

connecting the conductors across the terminals of a battery (figure 24.22).

In the electronic flash unit of cameras, chemical energy from a battery becomes converted into electrical energy that is stored in the capacitor. When the capacitor discharges, the electrical energy is converted into light energy by a special lamp tube in the camera flash unit.

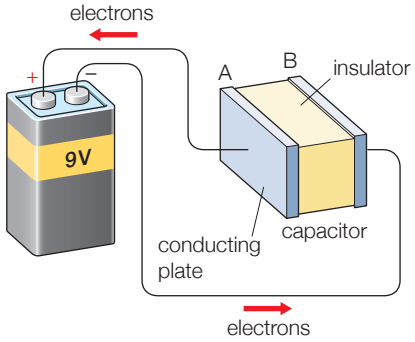


Figure 24.22 A battery being used to charge a simple capacitor. The arrows show the direction of movement of electrons through the leads during the charging process.

Capacitance

Depending on their construction, some capacitors can hold more charge per unit potential difference (see page 396 for definition) across the conducting plates than others. The **capacitance**, C, of a capacitor is defined as the charge stored per unit potential difference across the plates. Thus, if the charge stored is Q coulombs and the potential difference provided across the plates is V_c volts, then

$$C = \frac{Q}{V_c}$$

Capacitance is measured in farads (F), where one farad is equal to one coulomb per volt. One farad is a very large capacitance, which is difficult to obtain ordinarily so capacitance is normally quoted in a small unit, the microfarad (μF). One microfarad is 1/1 000 000 of a farad.

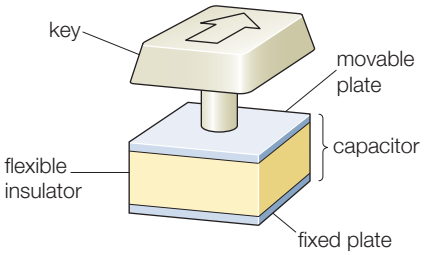


Figure 24.23 Computer keys.

capacitance ›

Keys on a computer keyboard are connected to a capacitor (figure 27.23). Pressing the key decreases the spacing between the 'plates', so increasing the capacitance. This is detected electronically.

IT Q8

A 2000 mF capacitor is charged to 20 V. How much charge does it store in the process?

The Van de Graaff machine

The Van de Graaff machine can store a great deal of charge. As it does so, it can achieve potential differences of up to a few million volts. The charge is obtained by friction between special moving belts and rollers inside the machine. The charge is collected by 'action of points' and induction on a dome resting on insulators. Both the large discharges and the high voltages have been used in conducting research, for instance into the structure of the atom.

A great deal of charge can be stored on the large metal dome since the gentle curvature reduces the possibility of charge leakage by 'action of points' and the rim of the dome is curved well inside the metal dome.

Chapter summary

- Charging usually involves adding or removing electrons from atoms.
- Objects can be charged by friction, induction or contact.
- Like charges repel and unlike charges attract. This behaviour can be used to explain why charged objects can even attract substances with zero net charge.
- In the atoms of conductors, the outermost electrons are few and are very weakly bound to the atoms.
- In the atoms of insulators, the outermost electrons are firmly bound to the atoms.
- Electric fields are regions in which electric charges experience electric forces.
- Electric fields are greatest near to point charges and fairly uniform between charged parallel conducting plates.

- Charge can be stored by capacitors or by a Van de Graaff machine.
- Capacitance is charge stored per unit potential difference in a capacitor ($C = Q/V$).
- Static electricity may be hazardous, as in lightning strikes or in sparking produced in the vicinity of combustible material.
- Useful applications, however, include dust precipitation, photocopying, spray painting and storage by capacitors for a variety of functions in electronics.

Answers to ITQs

ITQ1 It will repel it, since a plastic pen rubbed with cloth acquires a negative charge, and like charges repel.

ITQ2 The new distance is four times the original distance, so the final force is $1/4^2 = 1/16$ times the original force: $(1/16) \times 40\,\text{N} = 2.5\,\text{N}$.

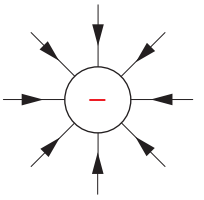
ITQ3 Since the polythene rod becomes negatively charged, it means that electrons were taken from the woollen cloth, leaving the cloth positively charged.

ITQ4 The negatively charged rod repels electrons as they approach the side of the ball that is nearer the rod.

ITQ5 If the rod, placed near to the cap of the electroscope, causes the leaf to diverge further, then the net charge on the rod is positive. (Note: If the divergence of the leaf decreases, we cannot be sure if the net charge on the rod is negative or zero. The *repulsion* test is the only sure test for the sign of a charge.)

ITQ6 A charged cloud overhead can induce a large charge on the point of the knife or scissors, resulting in 'action of points'. Charge could flow to or from the point, through someone holding the knife, resulting in electric shock

ITQ7



ITQ8 $Q = CV_c$
$= 2000 \times 10^{-6}\,\text{F} \times 20\,\text{V}$
$= 40\,000 \times 10^{-6}\,\text{C}$
$= 40\,000\,\text{mC}$

Examination-style questions

1 A Perspex pen rubbed with cloth becomes positively charged. Which of the following best explains this occurrence?

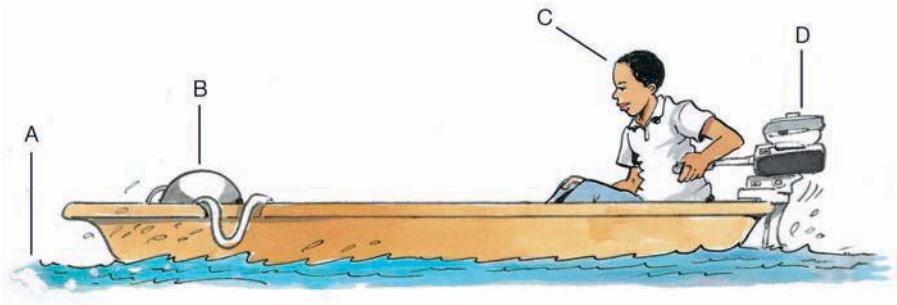
A The Perspex gains protons.

B The Perspex loses electrons.

C The cloth gains protons.

D The cloth loses electrons.

2 The diagram shows a fisherman in a small boat on the ocean during a thunderstorm. Which of the four points indicated is *most* likely to be struck by lightning from a charged overhead cloud?



3 A 600 μF capacitor is charged to 12V. How much charge is stored in the process?

A 20×10^{-6} C

B 72×10^{-6} C

C 50×10^{-6} C

D 7.2×10^{-3} C

4 A charged rod is brought near to a tiny piece of paper.

(i) The charged rod attracts the paper. What *two* possible inferences might be made about the paper?

(ii) The paper falls off the rod shortly after making contact with the rod.

Give an explanation for this.

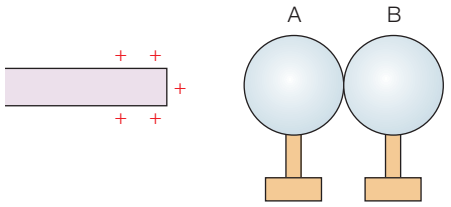
5 A negatively charged metal sphere, A, is connected by a copper wire to a similar sphere, B, that is positively charged. The initial charges on A and B are –5 C and +9 C, respectively.

(i) What kind of charge actually moves in the wire when the spheres are connected?

(ii) In which direction does the charge flow?

(iii) Estimate the size and sign of the final charge on each metal sphere.

6 A positively charged rod is brought close to two uncharged metal spheres A and B, which are in contact (see diagram).



The spheres are separated while the charged rod is held close to A. The rod is then removed.

(i) What are the signs of the charges left on A and B?

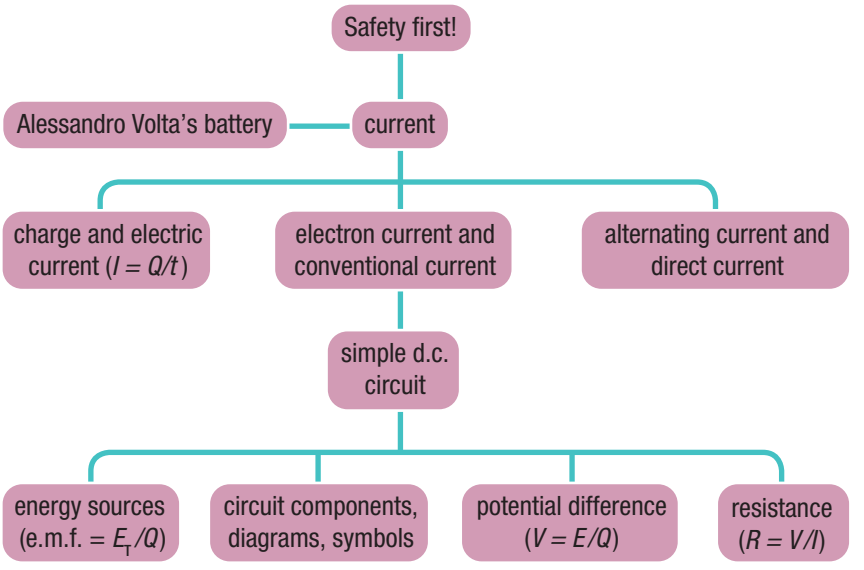
(ii) Explain the answer you have given in (i).

(iii) What is the name given to the charging process that has occurred?

25 Current Electricity

By the end of this chapter you should be able to:

- describe an electric current in terms of positive and negative charge carriers
- distinguish between electron flow and conventional current
- state the unit of electrical current, the ampere
- define the unit of electric charge, the coulomb, as one ampere second
- recall the relationship $Q = It$, and apply it to solve simple problems
- distinguish between alternating and direct currents
- cite examples of the conversion of electrical energy to other forms and vice versa
- define the potential difference between the ends of a conductor as the energy converted to other forms per unit charge flowing through the conductor, i.e. $V = E/Q$
- use the relationship $V = E/Q$ to solve problems involving energy transformation
- explain the concept of resistance
- determine resistance using the formula $R = V/I$
- use circuit symbols for cells, switches, wires, fuses, fixed and variable resistors, filament lamps, voltmeters, ammeters and semiconductors diodes
- draw simple circuit diagrams to represent a given arrangement of electrical components
- set up a simple circuit, given a circuit diagram, taking into account the polarity and suitability of components



Safety first

In chapter 24, we learned that a spark from static electricity near flammable material could start a fire. Electric shock can also occur, leading to death, if a certain amount of discharge current flows through the body.

BIOLOGY: muscle-contraction and electrical impulses

As well as static electricity, current electricity is also very dangerous (figure 25.1 shows the sign to look out for). Since the currents involved are much larger and longer-lasting than those usually found with static electricity, the possibility of shock increases. Muscles can contract so that a person is unable to let go of the source of electricity. The respiratory system can become paralysed. The heart can go into 'fibrillation' (irregular beating), which becomes difficult to stop, often leading to death.

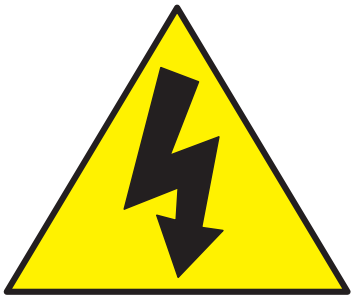


Figure 25.1 The international symbol warning of the presence of hazardous electricity.

When the skin is moist, the danger of current electricity to the body is heightened, since water containing dissolved salts is a conductor of electricity. Hence, electric circuits should never be handled with wet hands or wired near to sources or containers of water. Electrical fires should not be put out with water, but with extinguishing agents such as dry powder or carbon dioxide (figure 25.2).



Figure 25.2 A dry-powder fire extinguisher is used for electrical fires.

Heating always accompanies the passage of current electricity. So, we need to prevent excessive currents travelling along wires or through devices, to avoid the risk of fire caused by overheating.

Effects of electric currents can occur very quickly, increasing their danger. Paralysis of muscles can happen even before we are able to react. Current can destroy some electronic circuits in less than a millisecond (1/1000 second). Thus there is a need for extreme caution when dealing with current electricity.

In this course, most of the circuits use low voltages, and hence the currents are small and generally safe. However, by practising safe habits in low-voltage situations, we prepare ourselves for the safe handling of circuits involving higher voltages and currents. Better still, the safety habits can carry over into our daily lives in which electric current plays such a pervasive role.

A simple invention that revolutionised the world!

battery

When Alessandro Volta of Italy made public his newly invented '**battery**' in 1800 (figure 25.3), little did he realise that it would revolutionise people's lives worldwide.

The amazing feature of Volta's 'battery' was that it could maintain a relatively large, *steady* flow of charge, in other words, a current, for a long time. This was quite unlike the usually tiny, *momentary* flow produced during static discharge.



Figure 25.3 Volta exhibiting his 'battery' to Emperor Napoleon I in 1801.

The 'battery' current could also keep flowing around a circuit, whereas static discharge merely involved piling up the charge from one object to another.

Volta's invention of the battery opened the door to experiments with current electricity. Currents are now used everywhere – in homes, industry, communications, entertainment, business, medicine and even in space travel.Volta's 'battery' consisted of a series of 'cells' (figure 25.4). Each 'cell' consisted of a cup containing salt solution, or dilute acid, and two dissimilar metals such as silver and zinc. Another type of 'cell' consisted of cloth, soaked with the solution and sandwiched between the two metals shaped as discs.

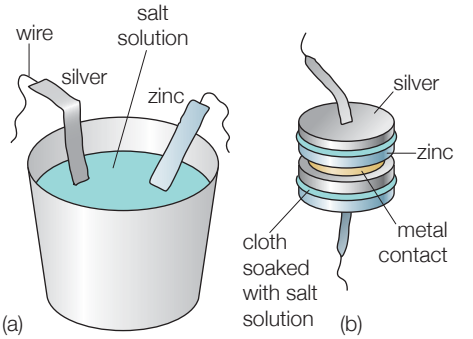


Figure 25.4 Volta's cells: (a) a cup cell; (b) a battery of two disc cells. Some modern cells and batteries are shown on pages 460–463.

Internet search terms: history of batteries/ Howstuffworks first battery

Electric current and charge

Charge carriers

electron ›

Electric current constitutes a flow of electric charge. In chapter 24 we saw that negative charge carriers, called **electrons**, move readily along conductors. The current in a conductor can therefore be called an electron current.

Ions, you will recall, are atoms or groups of atoms that have gained or lost electrons.

We saw also that the discharge current from a pointed conductor (page 378) involved ions as the charge carriers. The ionic current was a flow of either positive or negative charge through the air. When an electric current passes through a liquid, the charge carriers are also ions, and we will revisit this in chapter 26.

In chapter 30, we will learn of another kind of charge carrier called a 'hole'. A 'hole' is a vacancy left by an electron. 'Holes' are positive charge carriers and are characteristic of substances called semiconductors. Thus, we can speak of 'hole' currents.

Electron current and conventional current

In the early days of experiments with cells, scientists knew that there was a flow of charge in their circuits, and guessed that positive charge was flowing from the positive terminal of the battery, through the circuit, to the negative terminal. So they marked this as the current direction on their circuit diagrams. We now know that they guessed wrongly, and the current was actually a flow of negatively charged electrons in the opposite direction, but we still use their convention.

Electric current can involve a flow of electrons, ions or holes. By convention, we show the direction of an electric current as the direction of flow of *positive* charge (figure 25.5). Thus, electrons travelling in one direction are shown in diagrams of electrical circuits as a 'conventional current' travelling in the opposite direction. In fact, it is found that electrons travelling in one direction produce the same effects as positive charge travelling in the opposite direction.

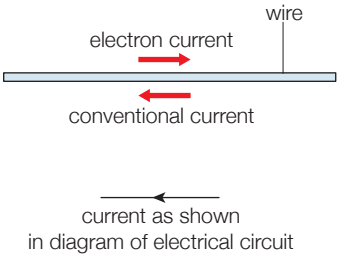


Figure 25.5 An electron current in one direction is treated as a conventional current in the opposite direction.

IT Q1

State the direction of conventional current for a current involving the following types of charge flowing from a point A to a point B. (The first one is done for you.) (i) Electrons [*Answer*: conventional current is from B to A]; (ii) positive ions; (iii) negative ions; (iv) holes.

The ampere and the coulomb

ampere ›

Experiments have shown that there is a force between two parallel wires when they are carrying a current. The S.I. unit of current, the **ampere** (abbreviated A), is defined in terms of this force.

One ampere is defined as that current which, in two long, thin parallel conductors, 1 m apart in a vacuum, produces a force of exactly 2×10^{-7} N per

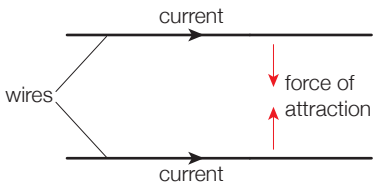


Figure 25.6 Wires carrying current exert a force on each other.

metre length of each conductors (figure 25.6). The coulomb (abbreviated C) is the S.I. unit of charge and is defined in terms of the ampere.

Current is defined as the *rate* of flow of charge past a particular point. If a quantity of charge of Q coulomb flows past a point in a time t seconds, the current I, in amperes, is given by

$$I = \frac{Q}{t}$$

and thus

$$Q = It$$

From the above equation, we see that

$$1\,\text{C} = 1\,\text{A} \times 1\,\text{s}$$

or

1 coulomb = 1 ampere second

CHEMISTRY: electrochemistry (the coulomb)

Hence, 1 coulomb of charge can be defined as the charge that flows in 1 second past a given point when the current is 1 ampere.

Worked example 25.1: Calculating charge

Charge from a current of 5 A flows for 10 s. How many coulombs of charge have flowed?

Solution

Current, $I = 5\,\text{A}$

Time, $t = 10\,\text{s}$

Charge, $Q = ?$

$$I = \frac{Q}{t}$$

Therefore

$$Q = It$$

$$= 5\,\text{A} \times 10\,\text{s}$$

$$= 50\,\text{C}$$

50 C of charge have flowed.

IT:Q2

If 16 C of charge flows past a point during 8 s, what is the average current?

Alternating and direct current

alternating current

The current that comes into our homes does not travel in one direction only. Electrons flow back and forth along each wire 60 (or 50) times each second. This type of current is called **alternating current** (a.c.). The frequency of an alternating current is the number of times electrons flow back and forth each second. The S.I. unit for frequency is the hertz (Hz). Thus, the current in our electricity supply has a frequency of 60 Hz (or 50 Hz in some regions).

direct current

In a battery current, however, charges flow in one direction only. This is called **direct current** (d.c.). We shall show voltage–time waveforms for alternating and direct currents in chapter 30.

Energy and electric circuits

The need for energy

circuit

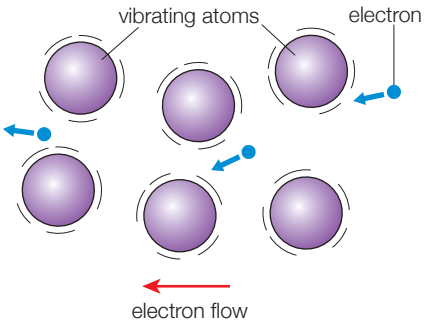
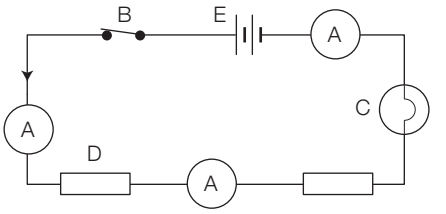


Figure 25.7 As electrons travel, they collide with vibrating atoms and lose energy. An energy source is required to keep the electrons moving in one direction.

A closed loop in which charges flow is called a **circuit**. It takes energy to keep electrons travelling around a circuit. This is because there are vibrating atoms present in every material. As the electrons travel, they collide with these atoms and lose energy to them (figure 25.7). The atoms, having gained energy, vibrate with greater speed and the material heats up as the average speed of the atoms increases (see page 173). Since the electrons are losing energy, they would eventually come to a stop, unless they were continually supplied with more energy. An energy source is therefore necessary to keep the electrons flowing in a circuit.

We can compare the flow of electrons round a circuit to the flow of water through pipes (figure 25.8). The water pump supplies the energy to keep the water circulating continuously. If the pump were to stop supplying energy, the flow of water would stop. Similarly, in an electric circuit a battery supplies energy to electrons to keep them flowing.



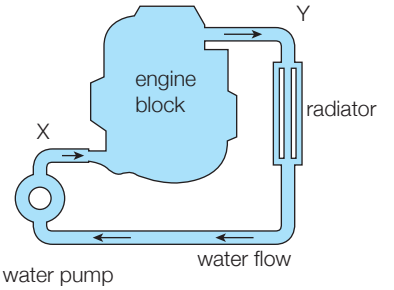


Figure 25.8 A water circuit in a water-cooled car engine.

Cells and batteries

cell

electrolyte

A **cell** can provide the energy required to keep electrons flowing in a circuit. A cell usually consists of two conductors made of dissimilar materials in contact with an **electrolyte** (figure 25.9). (Electrolytes are liquids through which a current of electricity can be carried by mobile ions.) When a device is connected across the conductors, electrons flow from one of the conductors, called the **negative terminal**, through the device to the other conductor or **positive terminal** (figure 25.10). Thus, a cell provides energy that enables electrons to flow through a device, and maintain a current in the circuit. A battery is a combination of cells.

negative terminal

positive terminal

CHEMISTRY: electrolytes

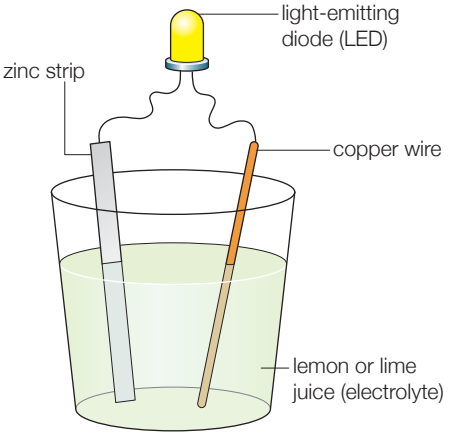


Figure 25.9 A simple cell you can make yourself. The cell can activate a light-emitting diode (LED; see page 477).

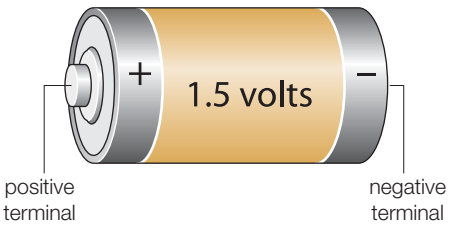


Figure 25.10 The positive and negative terminals marked on a common type of cell. The internal structure of such a cell (as well as other types) is shown on page 460.

ITQ3

To what desired form of energy is electrical energy converted in the following devices? (The first one is done for you.) (i) electric iron [*Answer*: electrical energy to heat energy]; (ii) loudspeaker; (iii) rechargeable battery during charging.

component ›

circuit diagram ›

A simple direct current (d.c.) circuit

When a cell or battery is used in a simple circuit, the electrons flow in one direction only; the cell provides a direct current. The circuit is therefore called a direct current (or d.c.) circuit.

A simple d.c. circuit consists of a d.c. energy source, such as a battery, a load that converts the electrical energy to a desired form, and connecting wire. A switch is used to control the current by either allowing or interrupting the electron flow. When the switch is open, the circuit is broken and no electrons flow. When the switch is closed, electrons flow around the circuit.

The battery, load, connecting wire and switch are called **components** of the circuit. Figure 25.11 shows a simple d.c. circuit consisting of a battery, a lamp (the load), a switch and connecting wire. Figure 25.11 (a) is a picture of the circuit, whereas figure 25.11 (b) is a **circuit diagram**, showing the same circuit drawn using electrical circuit symbols.

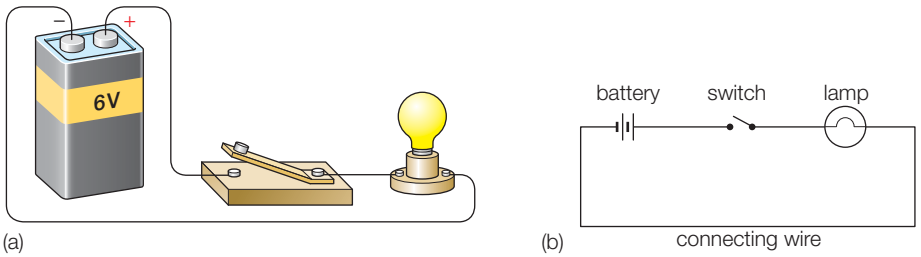


Figure 25.11 (a) A simple d.c. circuit. When the switch is closed, the lamp is converting electrical energy. (b) The same circuit shown as a circuit diagram.

Figure 25.12 shows the connections to a lamp bulb in more detail.

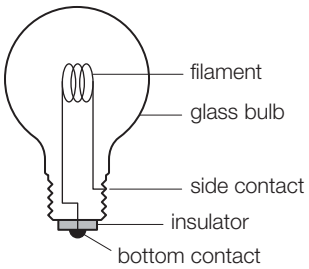


Figure 25.12 The connections for a lamp. One wire from the battery goes to the side contact; the other goes to the bottom contact.

Energy conversions

ELECTROCHEMISTRY: electrolysis

ITQ4

Figure 25.13 shows a flashlight. Will the bulb light with the slide switch, S, in the position shown? Explain your answer.

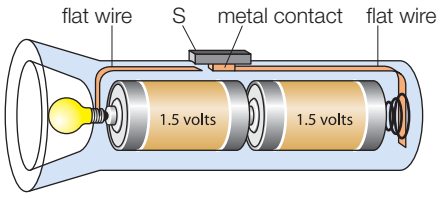


Figure 25.13 Will the flashlight light?

The main function of a lamp is to convert electrical energy into light energy. Other loads can perform different energy conversions. An electric motor converts electrical energy to mechanical (kinetic) energy. A buzzer converts electrical energy to sound energy. In an electrolytic cell, electrical energy is converted to chemical energy during the separation of chemical elements by electrolysis. Aluminium is extracted from aluminium oxide (bauxite) by electrolysis in such a cell.

Practical activity 25.1

Caution: To be done only under the direct supervision of your teacher.



Testing for good and poor conductors of electricity

(Very poor conductors of electricity are called insulators.)

You will need:

- a battery box containing two 1.5V zinc-carbon cells
- a switch
- a low-voltage lamp
- insulated connecting wire
- the materials to be tested (e.g. iron nail, wooden clothes pin, paper clip, plastic clothes pin, plastic spoon, metal spoon, 'lead' (actually carbon) inside a 'lead' pencil, tap water in a beaker, dilute salt water in a beaker, distilled water in a beaker, air).

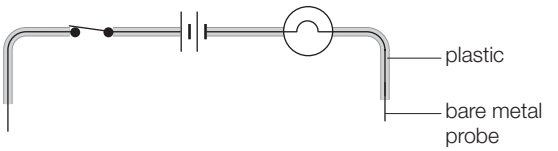


Figure 25.14 Testing materials for conduction of electricity.

Method

1 Clear your experiment area of objects, especially water.

2 Under the direct supervision of your teacher:

(a) Connect up the arrangement shown in figure 25.14, but with the switch in the open position and the battery outside the box.

(b) Insert the battery correctly and then turn on the switch.

(c) Holding the wires by the plastic, place the bare wire probes a suitable distance apart on, or in, the material being tested. Check if the bulb lights.

3 (a) Record your observations in an appropriate manner.

(d) Write a lab report, including a discussion as to whether the circuit can truly identify conductors and insulators, and stating a suitable conclusion.

CHEMISTRY: properties of metals and non-metals

Describing electrical circuits

Electromotive force (e.m.f.)

electromotive force (e.m.f.) ›

The **electromotive force (e.m.f.)** of a cell is the amount of energy it delivers per unit charge when it sends electrons around a complete circuit. If a cell delivers E_T joules of energy to the loads in sending a charge Q completely around a circuit, then:

$$\text{electromotive force, e.m.f.} = \frac{E_T}{Q} \text{ joules/coulomb}$$

volt ›

The unit of one joule per coulomb is called a **volt** and has the symbol V. Thus, e.m.f. is measured in volts, where $1\,V = 1\,J\,C^{-1}$. The 1.5 V stamped on a 'D' cell tells us the e.m.f. of that cell. A 9 V battery therefore delivers more energy per unit charge completely around a circuit than a 6 V battery.

Potential difference (p.d.)

potential difference ›

The **potential difference** or p.d., V, across a load in a circuit is the amount of energy being delivered to the load per unit charge following through it. Thus,

$$V = \frac{E_L}{Q_L} \text{ joules/coulomb}$$

The positive terminal of a cell is at a positive potential. The negative terminal is at a negative potential. The difference in electrical potential is the potential difference, called the e.m.f., of the cell, and causes charge of flow when the cell is connected in a circuit.

voltage ›

voltmeter ›

where E_L is the amount of energy delivered to the load when a charge Q_L passes through the load. Since potential difference is energy per unit charge (joule/coulomb), the volt is also the unit for potential difference. Another term used to represent potential difference is '**voltage**'.

Figure 25.15 shows how we connect a **voltmeter** to measure the potential difference across a lamp. With the lamp in a circuit, the voltmeter is connected *across* the two terminals of the lamp. The voltmeter is said to be **connected 'in parallel'** with the lamp.

parallel connection ›

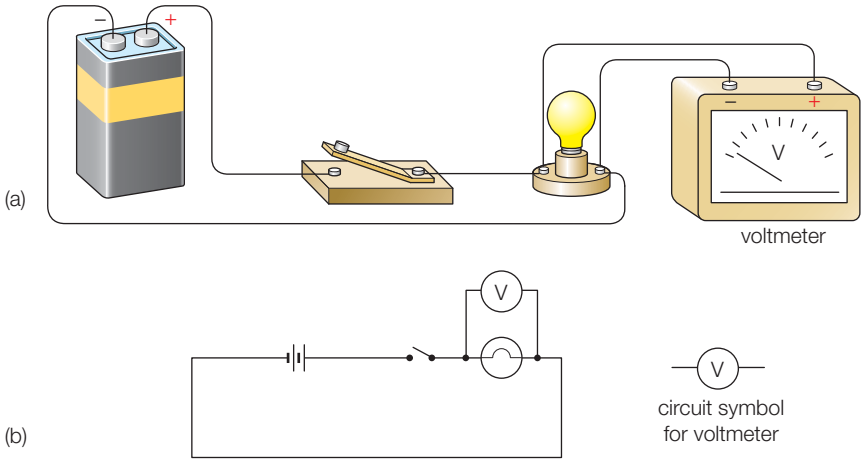


Figure 25.15 (a) Measuring potential difference across a lamp in a circuit. (b) The circuit diagram.

We define the volt as the potential difference between two points in a circuit if one joule of energy is needed to move one coulomb of charge between those points. P.d. represents a difference in electrical conditions (called potentials) between the two points.

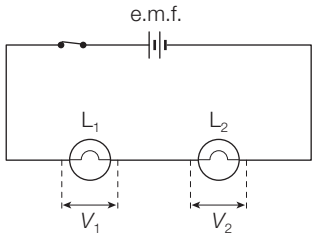


Figure 25.16 E.m.f. and potential differences in a series circuit.

The symbols V and V: In textbooks, a sloping (or italic) *V* is used to represent voltage or potential difference; and upright (or 'roman') V is the symbol for the volt as a unit.

Electromotive force and potential difference

Electromotive force (e.m.f.) is related to potential difference (p.d.).

Look at the circuit shown in figure 25.16. The circuit consists of only one loop; the components are connected in sequence. This is called a series circuit. The current in a series circuit can follow only one path. Thus current (charge flowing per second) is the same in each of the circuit components.

The e.m.f. is the total energy developed by the battery in transporting unit charge around the entire circuit. The p.d. across each component, or load, is the energy delivered per unit charge to that particular load. From the law of conservation of energy, the energy supplied by the battery must be equal to the total energy delivered to the loads. Thus, the e.m.f. in a circuit consisting of lamps L_1 and L_2 is the sum of the potential differences across L_1 and L_2. So

$$\text{e.m.f.} = V_1 + V_2$$

where V_1 and V_2 are the p.d.s across L_1 and L_2, respectively. (We are ignoring the small amount of energy used to transport the charge through the connecting wires and the materials of the battery itself.)

Worked example 25.2: Energy converted by lamp

How much energy is converted when 50 C of charge flows through a lamp that is connected to a 9.0 V battery?

Solution

Charge, $Q = 50$ C

Potential difference across lamp, $V = 9.0$ V

Energy converted, $E = ?$

Since $V = \frac{E}{Q}$

Therefore

$$E = QV = 50\text{ C} \times 9.0\text{ V} = 450\text{ J}$$

IT Q5

A lamp transforms 80 J of energy when a charge of 20 C flows through it. What is the potential difference applied across the lamp?

IT Q6

A 12 V battery is connected in series with two lamps, A and B. A voltmeter reads 5 V when placed across A. What reading would be expected if the voltmeter were placed across B?

Current

ammeter ›
series connection ›

We have already defined current, *I*, as the rate of flow of charge in a circuit (page 387). Figure 25.17 shows how we connect up an **ammeter** to measure current. Unlike the voltmeter, the ammeter is **connected in series** with

the load in the circuit. The circuit is first opened and the ammeter is inserted at the opening. The circuit does not have to be opened when a voltmeter is connected.

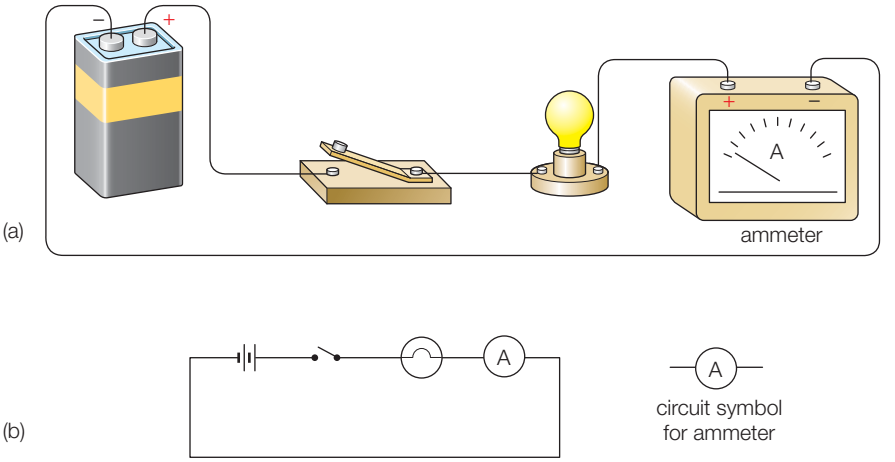


Figure 25.17 (a) Measuring the current passing through a lamp. (b) The circuit diagram.

Resistance

resistance ›

The **resistance** in a conductor is the opposition to the flow of electrons. This occurs because atoms are always vibrating. As noted earlier, when electrons flow they collide with these vibrating atoms, lose energy and slow down. The rate of flow of electrons (the current) around the entire circuit then decreases. Since the vibrations of atoms are more rapid at higher temperatures, collisions with electrons are then more frequent. Hence there is more opposition to electron flow, or greater resistance, at higher temperatures.

Georg Ohm, a German scientist, found that, for a given conductor kept at constant temperature, the potential difference, V, across the conductor is directly proportional to the current, I, flowing in the conductor, or

$V \propto I$ at constant temperature

Hence,

$V = I \times \text{constant}$, or $I = \frac{V}{\text{constant}}$

The constant is called the resistance, R, since, for a given potential difference, the equation indicates that the current, I, is small when the constant, R, is large. We usually write the equation as

potential difference = current × resistance

or

$$V = IR$$

Ohm's law ›
ohm ›
resistor ›

This equation is called the **Ohm's law** equation. The S.I. unit of resistance is named the **ohm** (Ω), after Georg Ohm.

Resistors are components used mainly to control the size of current flowing in circuits. Figure 25.18 shows an assortment of resistors and their corresponding electrical symbols. The resistance of a resistor depends on the material of which it is made and its dimensions.

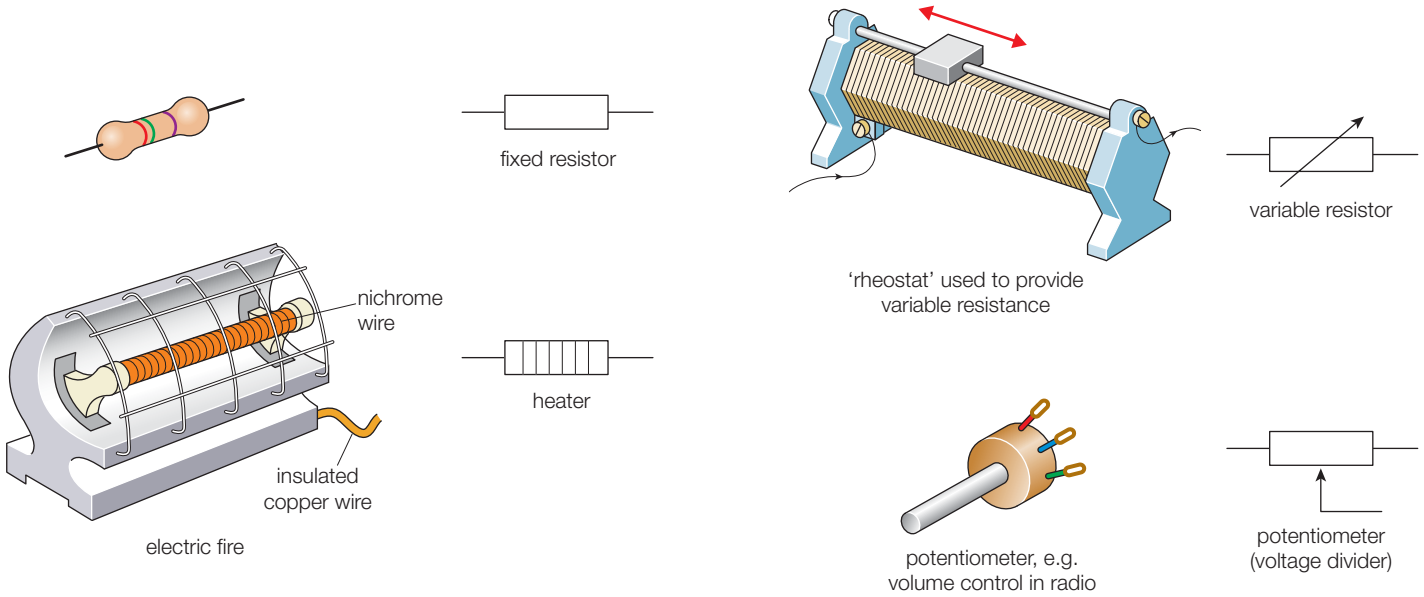


Figure 25.18 Types of resistors and their circuit symbols.

Semiconductor 'resistors' are discussed in chapter 30. Ohm's law does not apply to these 'resistors', however.

Worked example 25.3: Protecting a light-emitting diode (LED)

A particular light-emitting diode (LEDs are dealt with on page 477) will burn out if a current greater than 30 mA (30 milliamperes) passes through it. The LED is to be operated from a 6 V supply. How can the LED be protected from overheating?

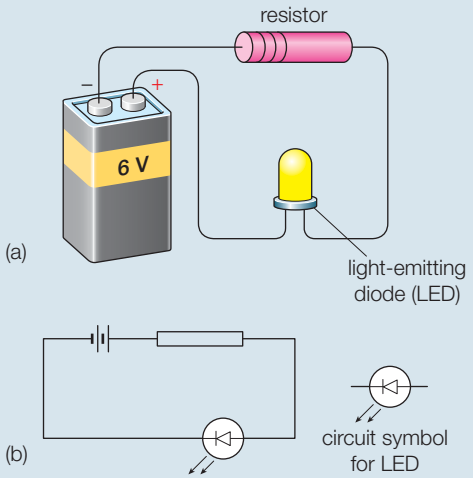


Figure 25.19 (a) Protecting a light-emitting diode (LED) using a series resistor. (b) Circuit diagram.

Solution

Connecting a resistor in series with the LED and the 6 V supply (figure 25.19) can solve the problem.

Ignoring any resistance in the battery, the wires and in the LED itself, we can calculate the necessary resistance, R, of the resistor using Ohm's law. Since $V = IR$, then

$$6 = 30 \times 10^{-3} \times R$$

Therefore

$$R = 200\,\Omega$$

A 200 Ω resistor in series with the 6 V battery will prevent the current from exceeding 30 mA.

Potential divider

Resistors are also used to obtain appropriate potential differences from a source of e.m.f. When used in this way, resistors are called potential dividers. (An example of using resistors as a potential divider is discussed on pages 401–402).

Short-circuits and overloads

fuse

fuse rating

Circuits are also protected using **fuses** (figure 25.20). The fuse wire is made of an alloy, usually of lead and tin. A fuse is connected in series with the battery. If the current exceeds a certain value, called the '**fuse rating**', the fuse wire overheats and melts. When the wire melts, the circuit breaks and current no longer flows.

In figure 25.21 (a), the total resistance of the circuit is 12 Ω, as the battery resistance is 1 Ω and the lamp resistance is 11 Ω. When the switch is closed, the lamp lights. Applying the Ohm's law equation, $V = IR$, we calculate that the current in the circuit is 0.5 A. This size of current does not cause overheating in the connecting wire. A fuse with a rating of 1 ampere will allow the 0.5 A current to flow and keep the lamp lit under normal conditions.



Figure 25.20 An assortment of fuses.

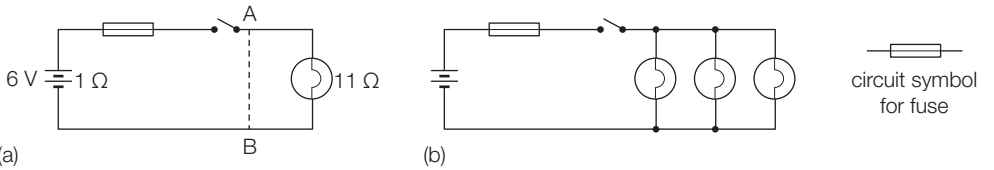


Figure 25.21 Circuit protection using a fuse. (a) Protecting for short-circuit. (b) Protecting for overload.

short-circuit

Supposing, however, that the wires were bare at A and B and they touched each other. A very large current would then pass through AB and hardly any through the lamp. This is because AB would be a path of much lower resistance than the lamp. We would call this a '**short-circuit**'. The resistance of the circuit is now only about 1 Ω (the battery resistance), since the filament resistance has been 'shorted-out' by AB. The current in this circuit would then be about 6 A (from $V = IR$). This size of current could make the wires and the battery overheat. The 1 ampere fuse, however, would 'blow' (melt) and break the circuit if there was a current of 6 A, since this current would exceed the rating of the fuse. This is an example of short-circuit protection.

Great care should be taken to match the current rating of connecting wires with the current expected in a circuit. A wire rated at 3 A will conduct up to 3 A of current without overheating. Thicker wires usually have higher current ratings than thin wires.

In figure 25.21 (b), if each lamp shown operates at 0.5 A, a total of 1.5 A of current will have to be delivered by the battery. A 3 A protection fuse might be appropriate. If lamps using higher currents (e.g. 2 A each) were used, the 3 A fuse would 'blow' since the total current would then be 6 A. Current would cease to flow and the wires would be protected from the overheating that would have been caused by the **overloaded** circuit.

overloaded circuit

Worked example 25.4: Choosing an appropriate fuse

A 12 V battery of resistance 1 Ω and connecting wire rated at 5 A are connected in series with a lamp of resistance 11 Ω. Which of the fuses shown in figure 25.20 should be used for short-circuit protection? Explain your answer.

Solution

Battery resistance, $R_1 = 1\ \Omega$

Lamp resistance, $R_2 = 11\ \Omega$

Since the resistances are in series, then if the total circuit resistance is R,

$$R = 1\,\Omega + 11\,\Omega$$
$$= 12\,\Omega$$

Battery e.m.f. = 12 V

e.m.f. = IR

Therefore current, I, is given by

$$I = \frac{\text{e.m.f.}}{R}$$
$$= \frac{12\,\text{V}}{12\,\Omega}$$
$$= 1\,\text{A}$$

Thus, for normal operation, I = 1 A.

In the case of a short-circuit, the 11 Ω resistance would be 'shorted' out and the total circuit resistance would be 1 Ω. Then the current, I, would be 12 V/1 Ω, or 12 A.

Since the wire rating is 5 A, then the 3 A fuse would be most appropriate, as it will allow the 1 A current for normal operation, but will blow before the current reaches 5 A. Thus, overheating of the wire and the battery will be prevented. A 5 A fuse is not suitable as it will allow currents dangerously too close to the wire rating of 5 A, beyond which overheating will take place.

From these examples we see that the choice of an appropriate fuse depends on the maximum current that the devices, as well as the connecting wires in a circuit, can safely take.

Setting up a d.c. electrical circuit

The circuit diagram

A circuit diagram is necessary when wiring electrical circuits. The circuit diagram helps us to arrange the circuit components correctly. We can also work out from the diagram what values of components should be used to prevent overheating, or what connections would produce short-circuits. The circuit diagram should first be approved by your teacher before any circuit wiring.

The circuit diagram must contain a switch. The switch is placed in series with the energy source (the battery). The switch is there to enable us to turn the current on or off when necessary.

The diagram must take into account the polarity (positive or negative) of the components. A battery has a positive and a negative terminal. Ammeters, voltmeters and some capacitors (the electrolytic type) also have terminals marked positive and negative. If the terminals marked positive on these devices are connected towards the negative terminal of a battery (or vice versa), these components can be damaged. The polarity of LEDs is also important, and they can be damaged or fail to light up if their terminals are not correctly connected. Before connecting up a circuit, it is good practice to trace, on the diagram, the (conventional) current from the positive terminal of the battery and ensure that the current enters at the positive terminal of each component in the circuit.

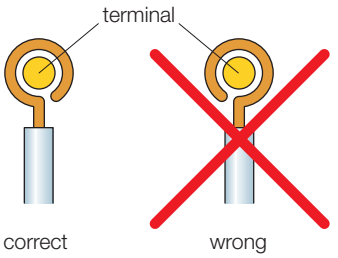


Figure 25.22 Passing connecting wire round a terminal.

Steps in wiring

1 Ensure that the area where you are wiring is dry and free of clutter.

2 Commence wiring, following the circuit diagram very closely. Do not connect any wire to the energy source (e.g. the battery) at this stage.

3 Make sure that all connections are clean and tight (but not too tight). Pass the connecting wire around a terminal in the same sense (normally clockwise) as that in which the terminal cap or nut is screwed on to the terminal (figure 25.22).

Use connecting wires that are insulated and whose ends are not frayed.

Be careful not to touch conducting parts of the circuit.

4 Finally, have your circuit wiring inspected by your teacher before you connect any wire to the battery.

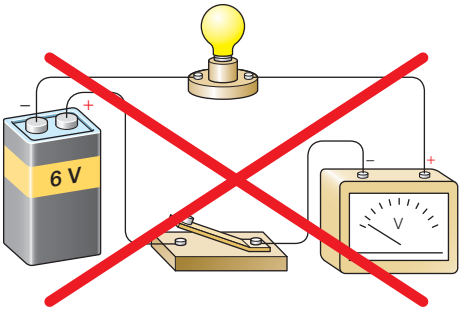


Figure 25.23 An incorrectly wired circuit.

Practices like these assist in the safe use of current electricity.

ITQ7

In what two ways is the voltmeter in figure 25.23 incorrectly connected for measuring the p.d. across the lamp?

Practical activity 25.2

Making and investigating battery cells

Caution: Electricity



There has been much interest in making better and better battery cells in recent times. This is because many countries are encouraging the development of battery-operated cars. You are to construct and test a cell using two different conductors and an electrolyte. (For safety reasons, do not use strong acids, such as sulphuric, hydrochloric or nitric acid. You can use a weak acid, such as found in fruit juices, a salt solution or even coconut water as the electrolyte.) Design and carry out an investigation to find out *one* of the following aspects (or any other aspect that interests you). You will need to pay strict attention to control of variables.

1 The effect of the different types of conductors on voltage obtained.

2 The effect of the surface area of the conductors on the current obtained.

3 Discharge characteristic (voltage vs. time) for your cell (and possible uses).

4 The 'life' of two types of cells compared (e.g. how long each cell can power a light-emitting diode).

Internet search term: electric car batteries

Chapter summary

- Current is defined as the rate of flow of charge. The current I associated with a charge of Q that flows during a time t is given by:
  $I = \frac{Q}{t}$
- The S.I. unit for current, the ampere, is defined in terms of the force between two parallel wires carrying a current.
- The S.I. unit for charge, the coulomb, is defined by the equation
  $Q = It$
  where $I = 1$ A and $t = 1$ s.

- Charge carriers may be positive or negative. Electrons, ions and holes are examples of charge carriers in a current.
- The equivalent current of positive charge in a circuit is called conventional current.
- In direct current (d.c.), charge carriers flow in one direction only.
- In alternating current (a.c.), the charge carriers move back and forth along the same wire.
- The potential difference between the ends of a conductor is the difference between their electrical conditions (or potentials).
- The potential difference, V, between the ends of a conductor is defined as the energy converted per unit charge flowing through it, Thus

  $V = \frac{E}{Q}$

  where E is the energy converted when a charge Q flows through the conductor.
- Resistance is the opposition to the flow of charge carriers. Resistance, R, is defined by the Ohm's law equation:

  $V = IR$, or $R = \frac{V}{I}$
- We use circuit symbols when drawing circuit diagrams.
- When designing and wiring electrical circuits, attention is paid to the polarity of components and to the current rating of wires and fuses.

Answers to ITQs

ITQ1 **(i)** B to A; **(ii)** A to B; **(iii)** B to A; **(iv)** A to B.

ITQ2 $I = Q/t = 16\,\text{C}/8\,\text{s} = 2\,\text{A}$

ITQ3 **(i)** Electrical energy to heat energy; **(ii)** electrical energy to sound energy; **(iii)** electrical energy to chemical potential energy.

ITQ4 No, the bulb will not light because there is a gap (an open circuit) at the switch.

ITQ5 $V = E/Q = 80\,\text{J}/20\,\text{C} = 4\,\text{V}$

ITQ6 If V_A and V_B are the p.d.s across lamps A and B, then $V_A + V_B = 12\,\text{V}$; therefore $5 + V_B = 12\,\text{V}$, or $V_B = 7\,\text{V}$

ITQ7 The voltmeter should be connected in parallel with L_1 and not in series with it (see figure 25.15). The positive terminal on the voltmeter should be connected towards the positive side of the circuit.

Examination-style questions

1 Which of the following can only carry a positive charge?

A electrons B holes C ions D neutrons

2 A charge of 20 C flows for 5.0 s. The current is:

A 0.25 A B 4.0 A C 15 A D 25 A

3 How much energy is converted when 3 C of charge is moved through a potential difference of 12 V along a wire?

A 0.25 J B 4.0 J C 9.0 J D 36 J

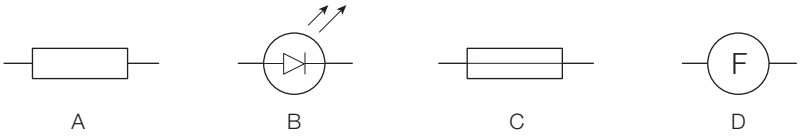
4 Which of the following units is used to measure charge?

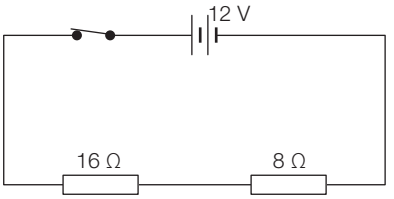
A ampere (A) B coulomb (C) C joule (J) D volt (V)

5 What is the *first* thing that happens when an excessive current passes through a fuse?

A The current is cut off.
B The circuit breaks.
C The fuse wire gets hot.
D The fuse wire melts.

6 Which of the symbols below represents a fuse?





7 The diagram on the left shows an electric circuit.

The circuit can *best* be described as

A overloaded
B parallel
C series
D a short-circuit

8 The circuit on the right shows an arrangement in which the brightness of a lamp can be varied, and certain measurements taken.

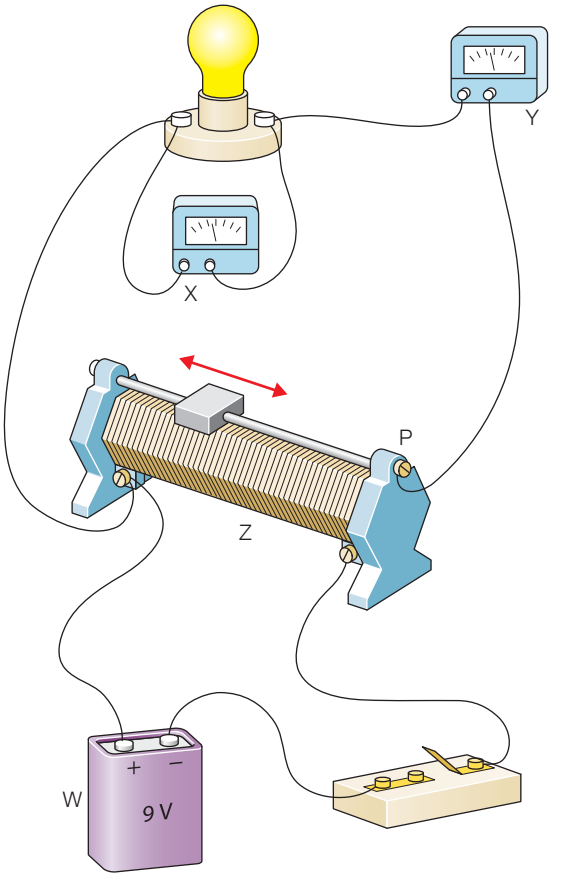
(i) Draw a circuit diagram to represent the above arrangement of components.

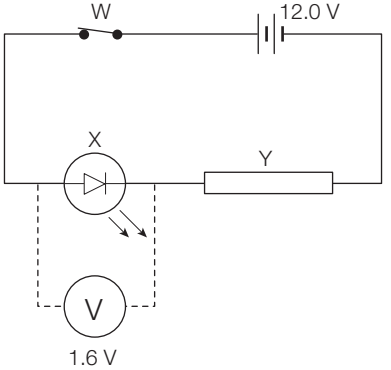
(ii) Name the components labelled W, X, Y and Z.

(iii) Give the special name of the arrangement of Z and W that provides the voltage to operate the lamp.

(iv) If the current in the lamp is 1.5 A when the potential difference across the lamp is 6.0 V, calculate the resistance of the lamp.

(v) What would be the effect on the brightness of the lamp if the slider were moved towards P? Explain your answer.





9 The diagram shows a series circuit. A voltmeter connected across the device, X, reads 1.6 V.

(i) Name the devices W, X and Y.

(ii) What is the function of Y?

(iii) What would the voltmeter read if placed across Y?

(iv) If a maximum current of 0.020 A is to flow in the circuit, determine the value of Y.

26 More about Electrical Circuits

By the end of this chapter you should be able to:

- differentiate between series and parallel circuits
- recall that the current in a series circuit is the same everywhere in the circuit and use this principle to solve problems
- recall that the net potential difference across any number of components connected in series is equal to the sum of the potential differences across the components (i.e. $V_s = V_1 + V_2 + V_3 + ...$) and apply this to solve problems
- recall and use the following formulas for finding the total effective resistances
  $R_s = R_1 + R_2 + R_3 + ...$ for resistors in series
  $\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + ...$ for resistors in parallel
- recall that the sum of the currents in the branches of a parallel circuit is equal to the current entering or leaving the parallel section (i.e. $I = I_1 + I_2 + I_3 + ...$) and apply this concept to circuit problems
- recall that the potential differences across any number of components in parallel are the same (i.e. $V_p = V_1 = V_2 = V_3 = ...$) and apply this concept to solve problems
- explain why an ammeter should have a very low resistance
- explain why a voltmeter should have a very high resistance
- describe experiments using an ammeter and a voltmeter to investigate the relationship between current, *I*, and potential difference, *V*, for
  - metallic conductors at constant temperature
  - filament lamps
  - semiconductor diodes
  - solutions of copper sulphate in water using copper electrodes
- draw *I*–*V* graphs from the result of such experiments and draw appropriate conclusions from the graphs obtained
- recall that power $P = VI$, and use this relationship to solve problems

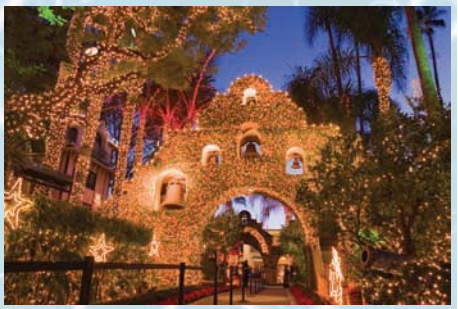
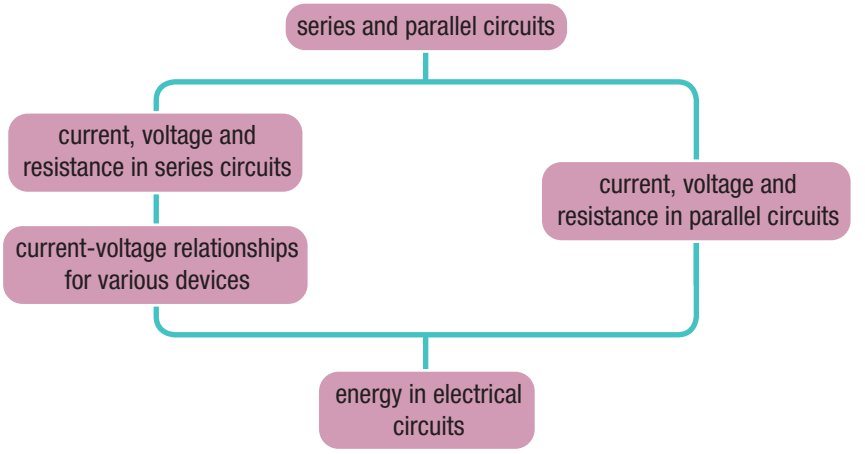


Figure 26.1 Diwali lights. But how many strings of lights could safely be connected to one voltage supply power cord? The answer is in this chapter.



Introduction

Modern society celebrates festivals such as Christmas and Diwali (figure 26.1) with strings of glowing electric lights. Society today is also making increasing use of electrical equipment and appliances. This chapter considers two types of electrical circuits – series and parallel – in which electricity is employed for lighting and other uses. It also considers some basics about energy and power in these circuits.

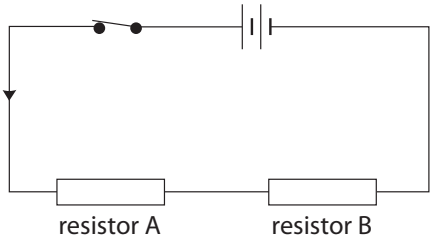


Figure 26.2 The components of the circuit are all connected in series.

Series circuits

series circuit ›

A **series circuit** consists of components connected in sequence, as in figure 26.2.

Current in series circuits

Current takes a *single* path in a series circuit.

Figure 26.3 (a) shows ammeters placed at three different places in a series circuit to measure electric current. The flow of the electric charge is in some ways like water flowing in the cooling system of a car. In the water circuit shown in figure 26.3 (b), the rate at which water leaves the pump is the same as the rate at which it returns to the pump. The question arises: Is the rate of flow of electrons (i.e. the current) the same throughout the circuit?

ITQ1

What do the symbols B, C, D, E and F in figure 26.3 (a) represent? (Hint: see chapter 25.)

ITQ2

In figure 26.3 (b), is the rate of flow water at X, which is nearer the pump, greater than that at Y? Explain your answer.

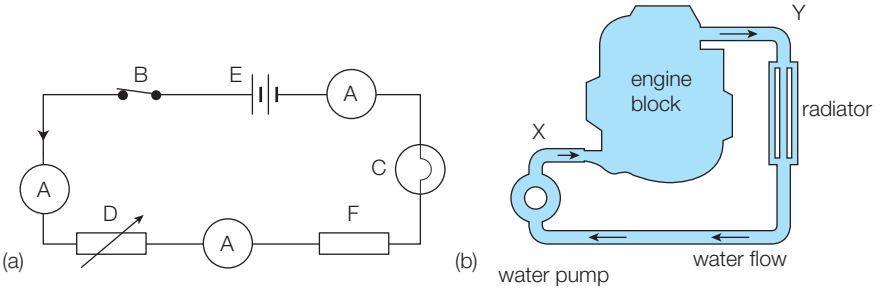


Figure 26.3 (a) Electrical circuit. (b) Water circuit in a motor car engine.

Practical activity 26.1

Current in a series circuit

Caution: To be done only under the direct supervision of your teacher.



You will need:

- a battery box containing two 1.5V zinc-carbon cells
- a switch
- a low-voltage lamp
- insulated connecting wire
- a variable resistor
- a fixed resistor
- three ammeters.

Method

1. Connect up the circuit of figure 26.3 (a). (Since you will be dealing with electricity, first refer to the last section of chapter 25, 'Setting up a d.c. circuit'.)
2. Vary the resistance of D and obtain and record about five sets of readings on the ammeters.
3. Write a suitable conclusion based on your results.

Question

Is the current the same at different points in a series circuit?

Experiments have shown that:

The current is the same throughout a series circuit.

Voltage (potential difference, or p.d.) in series circuits

Potential difference (p.d.) and voltage drop (or voltage) are terms that mean the same thing.

Figure 26.4 shows voltmeters connected across two resistors, A and B, separately and across A and B connected together in series.

Practical activity 26.2

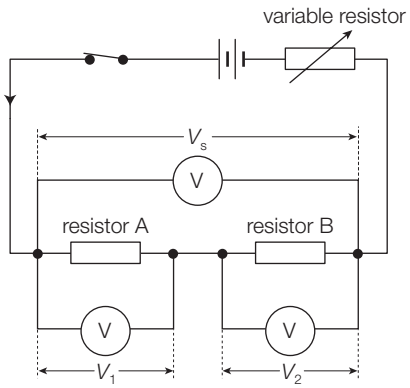


Figure 26.4 Measuring voltages in a series circuit.

Voltages in a series circuit

Caution: To be done only under the direct supervision of your teacher.



You will need:

- a battery box containing two 1.5V zinc-carbon cells
- a switch
- insulated connecting wire
- a variable resistor
- a fixed resistor
- three voltmeters.

Method

1. Connect up the circuit of figure 26.4. (Since you will be dealing with electricity, first refer to the last section of chapter 25, 'Setting up a d.c. circuit'.)
2. Vary the resistance of the variable resistor and obtain and record about five sets of readings on the voltmeters.
3. Write a suitable conclusion based on your results.

Question

What is the relationship between the voltage across both resistances A and B and the voltages across each of A and B?

With reference to figure 26.4, experiments have shown that $V_s = V_1 + V_2$. In other words:

The total potential difference across resistors connected in series is the sum of the potential differences across the individual resistors.

Resistance in series circuits

Look again at figure 26.4. The current, I, through resistor A and resistor B is the same. Since

$$V_s = V_1 + V_2$$

then, by Ohm's law (page 392),

$$IR_s = IR_1 + IR_2$$

$$R_s = R_1 + R_2$$

where:

- R_1 = resistance of resistor A;
- R_2 = resistance of resistor B;
- R_s = resistance of the series combination of resistors A and B.

The total resistance in a series circuit is the sum of the separate resistances.

Worked example 26.1: Resistors in series

Three resistors are connected in a series circuit as shown in figure 26.5.

Calculate:

(i) the total effective resistance of the resistor combination;

(ii) the p.d. across the 4.0 Ω resistor;

(iii) the p.d. across the 5.0 Ω and 6.0 Ω series combination.

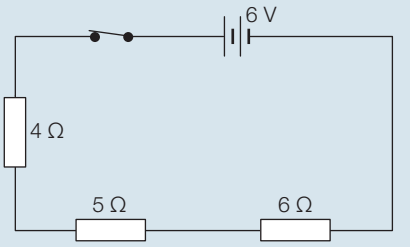


Figure 26.5

Solution

Let $R_1 = 4.0\,\Omega$, $R_2 = 5.0\,\Omega$ and $R_3 = 6.0\,\Omega$

Let R_s = total effective series resistance.

Let V = total p.d. across all three resistors in series.

(i) We have

$$R_s = R_1 + R_2 + R_3$$
$$= 4.0 + 5.0 + 6.0\,\Omega$$
$$= 15.0\,\Omega$$

(ii) To find the p.d., V_1, across the 4.0 Ω resistor, we must find the current through it. As this is a series circuit, this will be the same as the current, I, throughout the circuit.

$$V = IR_s$$

Therefore

$$I = \frac{V}{R_s} = \frac{6.0\,\text{V}}{15\,\text{W}} = 0.40\,\text{A}$$

Then, calculating the p.d. across the 4 Ω resistor,

$$V_1 = IR_1 = 0.40\,\text{A} \times 4.0\,\Omega = 1.6\,\text{V}$$

(iii) Let V_{2+3} = p.d. across the R_2 and R_3 series combination. Then

$$V = V_1 + V_{2+3}$$

and

$$V_{2+3} = V - V_1 = 6.0\,\text{V} - 1.6\,\text{V} = 3.4\,\text{V}$$

ITQ3

A string of 20 fairy lights, connected in series, is plugged into a 120 V mains circuit (figure 26.6). (i) What is the voltage drop across each bulb in the string? (ii) If the resistance of each bulb is 5 Ω, what is the current in the lamps?

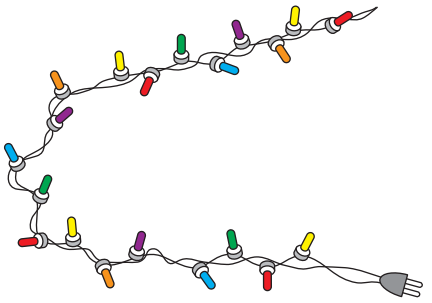


Figure 26.6 A string of 20 fairy lights.

potential divider ›

Since, for resistors connected in series, $V_s = V_1 + V_2$ and $R_s = R_1 + R_2$, we can use such a combination to obtain a small voltage from a larger voltage. The arrangement is called a **potential divider** and is described in Worked example 26.2 below.

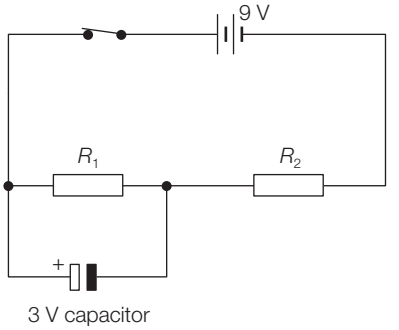


Figure 26.7 Two resistors forming a potential divider (voltage divider).

Worked example 26.2: Obtaining a low potential difference from a d.c. source – the potential divider

A capacitor requires a p.d. of 3 V for charging. How can we get such a p.d. from a 9 V battery?

Solution

We connect two resistors, R_1 and R_2, in series with the battery (figure 26.7). The e.m.f. of the battery provides a potential difference across the series combination. If we choose the correct ratio for the values of R_1 and R_2, we can obtain a p.d. of 3 V (V_1) across R_1 and the remaining 6 V voltage drop (V_2) across R_2.

Suppose we let $R_1 + R_2 = 180\,\Omega$ (an arbitrary value chosen because it is divisible by 3, 6 and 9). Then the current, I, applying Ohm's law, is given by

$$I = \frac{9\,\text{V}}{180\,\Omega} = 0.05\,\text{A}$$

To obtain 3 V across R_1, we must have $3 = IR_1$. That is,

$$3 = 0.05R_1$$

and so

$$R_1 = 60\,\Omega$$

Since $R_1 + R_2 = 180\,\Omega$, then $R_2 = 120\,\Omega$

The p.d. across R_2 will be 9 V – 3 V = 6 V.

Thus, a 60 Ω and a 120 Ω resistance in series divide a potential difference of 9 V in a ratio of 3 V : 6 V. This shows that $R_1/R_2 = V_1/V_2$.

This example of the use of a potential divider shows that if we need ⅓ of a given voltage, we should use ⅓ of the total resistance that is connected across the input voltage. This is in keeping with Ohm's law, that is, V is directly proportional to R.

Parallel circuits

Currents and voltages in parallel circuits

parallel circuit ›

Figure 26.8 shows resistors A and B connected in parallel. In a **parallel** arrangement, the current can take more than one path.

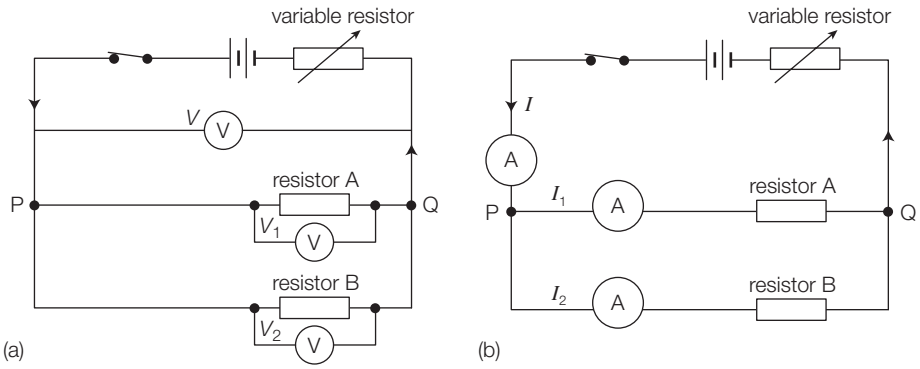


Figure 26.8 (a) Resistor A and resistor B are connected in parallel. (b) Measuring current in a parallel circuit.

Practical activity 26.3

Currents and voltages in a parallel circuit

Caution: To be done only under the direct supervision of your teacher.



You will need:

- a battery box containing two 1.5V zinc-carbon cells
- a switch
- insulated connecting wire
- a variable resistor
- two fixed resistors
- two voltmeters
- three ammeters.

Method

1 Connect up, one at a time, the circuits of figure 26.8. (Since you will be dealing with electricity, first refer to the last section of chapter 25, 'Setting up a d.c. circuit'.)

2 (a) Vary the resistance of the variable resistor and obtain and record about five sets of readings on the voltmeters across resistors A and B, and across the parallel combination of A and B (figure 26.8 (a)).

(b) Vary the resistance of the variable resistor and obtain and record about five sets of currents I, I_1 and I_2 (figure 26.8 (b))

3 Write a suitable conclusion based on your results.

Questions

1 What is the relationship between the voltages V_1 and V_2 across each of the resistances A and B and the voltage, V, across the parallel combination of A and B?

2 What is the relationship between the current, I, entering P and the currents I_1 and I_2 leaving P?

Experiments have shown that:

(a) The potential differences across resistors connected in parallel are equal, i.e. $V_1 = V_2$. Since both resistors A and B in figure 26.8 are connected to the same junctions P and Q, the potentials at P and Q are the same for both A and B, and so the voltage drop across them will be the same. It can be shown that the potential difference, V, across any number of resistors connected in parallel is equal to the potential difference across each resistor, i.e.

$$V = V_1 = V_2 = V_3 = \ldots$$

(b) The current splits up in a parallel circuit such that the total current entering a junction, such as P, is equal to the total current leaving the junction i.e.

$$I = I_1 + I_2$$

Another way of describing this is to say that the algebraic sum of the currents at any junction, such as P or Q, is zero. In other words, at P, for example, $I - (I_1 + I_2) = 0$. (Note that we represent current entering a junction as positive and current leaving the junction as negative.)

IT Q4

In figure 26.8, if the total current delivered by the battery is 1.4 A, and the current through resistor A is 0.8 A, what is the current through resistor B?

Worked example 26.3: Fairy lights

How many strings of fairy lights (figure 26.9), each using 1.2 A, can safely be connected in parallel to an extension power cord that has a rating of 3.0 A? (A rating of 3.0 A means that it can carry 3.0 A of current without overheating.)

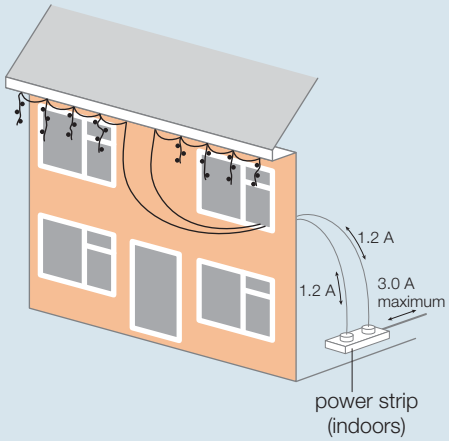


Figure 26.9 Strings of fairy lights connected in parallel to a power strip attached to a 3.0 A extension cord.

Solution

For strings in parallel, the total current delivered is equal to the sum of the currents in each string.

Since 1.2 A passes through each string of lights, then the maximum number of strings that should be connected to the extension cord is 3.10/1.2, or 2.5. Hence, no more than two strings could safely be connected in this situation.

Internet search term: fairy lights images

(These photos will show beautiful lighting arrangements made possible by strings of electric lamps)

Resistance in parallel circuits

Look again at figure 26.8. What is the effective resistance, R_p, of the parallel combination of the two resistors?

We know that

$$I = I_1 + I_2$$

Suppose the voltage across PQ is V. Then, applying Ohm's law

$$\text{current} = \frac{\text{p.d.}}{\text{resistance}}$$

to the above equation, we get

$$\frac{V}{R_p} = \frac{V}{R_1} + \frac{V}{R_2}$$

where:

- R_p = effective resistance of resistor A and resistor B connected in parallel;
- R_1 = resistance of resistor A;
- R_2 = resistance of resistor B.

Simplifying,

$$\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2}$$

This equation can be rearranged as follows to obtain a formula for determining R_p:

$$\frac{1}{R_p} = \frac{R_1 + R_2}{R_1 R_2}$$

Therefore

$$R_p = \frac{R_1 R_2}{R_1 + R_2}$$

product/sum formula

This last equation is known as the '**product/sum formula**' for finding R_p. Note that this formula can only be used to find the effective resistance of *two* resistors connected in parallel.

Combined series and parallel circuits

Resistance in circuits with both series and parallel connections

Worked example 26.4: Resistance in a series–parallel combination

Determine the total resistance, R_T of the circuit shown in figure 26.10.

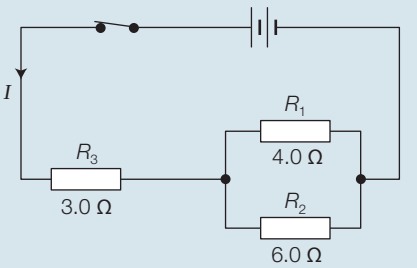


Figure 26.10 Series and parallel combination of resistors.

Solution

We first calculate the effective resistance, R_p, of the parallel combination as follows:

$$\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2}$$

$$\frac{1}{R_p} = \frac{1}{4.0} + \frac{1}{6.0}$$

$$\frac{1}{R_p} = \frac{5.0}{12}$$

So

$$R_p = \frac{12}{5.0} = 2.4\,\Omega$$

Or, using the produce/sum formula, we get

$$R_p = \frac{4.0 \times 6.0}{4.0 + 6.0} = \frac{24}{10} = 2.4\,\Omega$$

The effective resistance R_p is in series with R_3. Hence, the total circuit resistance, R_T, will be given by

$$R_T = R_p + R_3$$

$$R_T = 2.4 + 3.0$$

$$R_T = 5.4\,\Omega$$

In Worked example 26.4 you may have noticed that the effective resistance of the 4.0 Ω and 6.0 Ω resistors in parallel was only 2.4 Ω. The effective resistance after resistors are connected in parallel is always less than that of the lowest resistance. This makes sense, as the current now has more 'routes' through which it can flow.

Current in circuits with both series and parallel connections

Worked example 26.5: Current in a series–parallel combination

In the circuit shown in figure 26.10, determine total current, I, and the currents I_1 and I_2 through resistances R_1 and R_2, respectively, assuming that the e.m.f. of the battery is 6.0 V.

Solution

We use Ohm's law to calculate the total current, I. Since $V = IR$, then

$$I = \frac{V}{R}$$

$$I = \frac{6.0\,\text{V}}{5.4\,\Omega}$$

$$= 1.11\,\text{A or } 1.1\,\text{A (to 2 sig. fig.)}$$

To find I_1, the current in R_1, we must first find the p.d., V_1, across R_1. The p.d., V_3, across R_3 is given by

$$V_3 = IR_3$$

For simplicity, we are neglecting the resistance of the battery in this calculation.

$V_3 = 1.11\,\text{A} \times 3.0\,\Omega$

$V_3 = 3.3\,\text{V}$ (to 2 sig. fig.)

Since the battery e.m.f. is 6.0 V, the p.d., V_p, across the parallel combination is given by

$V_p = 6.0\,\text{V} - V_3$

$V_p = 6.0\,\text{V} - 3.3\,\text{V}$

$V_p = 2.7\,\text{V}$

But the p.d. across the parallel combination is the same as the p.d. across R_1 (as well as across R_2). So, V_2 is also 2.7 V. Since, by Ohm's law,

$I_1 = \frac{V_1}{R_1}$

then

$I_1 = \frac{2.7\,\text{V}}{4.0\,\Omega}$

$I_1 = 0.68\,\text{A}$ (to 2 sig. fig.)

We can now find I_2, the current through R_2. Since

$I = I_1 + I_2$

then

$I - I_1 = I_2$

$1.11 - 0.68 = I_2$

$I_2 = 0.43\,\text{A}$ (to 2 sig. fig.)

Current–voltage relationships

Different types of components behave differently when voltages are applied to them. To investigate the current–voltage relationship for a device, we use ammeters and voltmeters to measure the current and voltage, respectively.

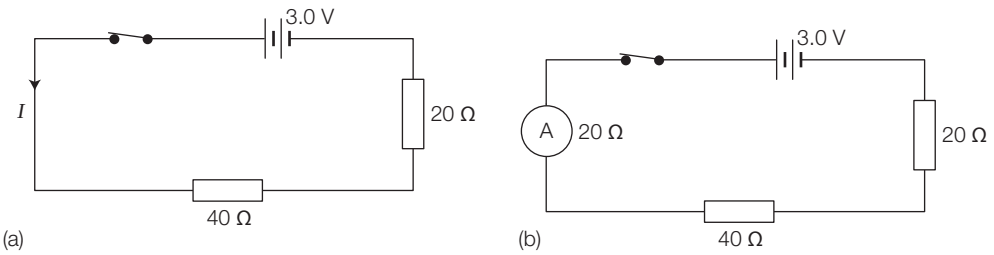


Figure 26.11 Using an ammeter. (a) simple d.c. series circuit with a 40 Ω and a 20 Ω resistor. (b) Ammeter inserted in the series circuit.

Using ammeters

Figure 26.11 (a) shows a simple circuit in which the total series resistance is 60 Ω. Applying Ohm's law (and neglecting the resistance of the battery and wires), we can calculate a value for the current:

$I = \frac{3.0\,\text{V}}{6.0\,\Omega} = 0.050\,\text{A}$

Suppose we wished to measure the current using an ammeter of resistance 20 Ω (figure 26.11 (b)). The total circuit resistance would be 80 Ω. The current I, would then be given by

$I = \frac{3.0\,\text{V}}{80\,\Omega} = 0.038\,\text{A}$ (to 2 sig. fig.)

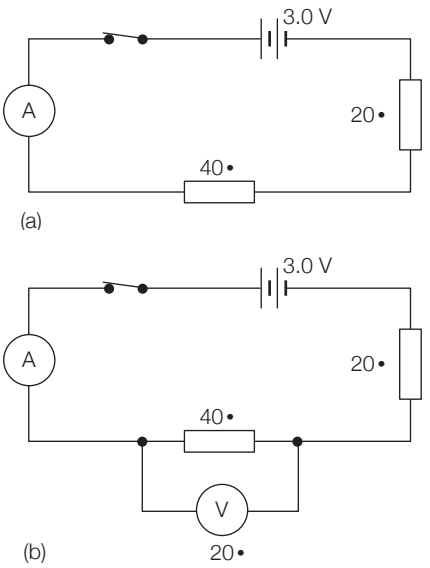


Figure 26.12 Measuring the potential difference across the 40 Ω resistor.

Inserting the ammeter reduced the current because the ammeter itself added resistance to the circuit. It is therefore important to use an ammeter of as low resistance as possible when attempting to measure current in a circuit.

An ammeter should have a very low resistance.

Using voltmeters

We can calculate the voltage drop across the 40 Ω resistor in the circuit shown in figure 26.11 (a). Since the current is 0.050 A, then from Ohm's law, the potential difference, V, across the 40 Ω resistor is 0.050 A × 40 Ω, i.e. $V = 2.0$ V.

Suppose the p.d. across the 40 Ω resistor was measured using a voltmeter of resistance 20 Ω, as shown in figure 26.12 (b). What would be the reading on the voltmeter?

To find this, we first calculate the resistance, R_p, of the parallel combination of the meter and the 40 Ω resistor:

$$\frac{1}{R_p} = \frac{1}{20} + \frac{1}{40}$$

$$\frac{1}{R_p} = \frac{2+1}{40}$$

$$R_p = \frac{40}{3} = 13.3\,\Omega = 13\,\Omega \text{ (to 2 sig. fig.)}$$

A quicker approach is to use the product/sum formula (page 405):

$$R_p = \frac{20 \times 40}{20 + 40} = 13\,\Omega \text{ (to 2 sig. fig.)}$$

The total resistance, R_p, of the circuit is now

$$13\,\Omega + 20\,\Omega = 33\,\Omega$$

The current, I, delivered by the battery is therefore given by

$$I = \frac{3.0\,\text{V}}{33\,\Omega} = 0.091\,\text{A} \text{ (to 2 sig. fig.)}$$

The reading, V, on the voltmeter will be the same as the potential difference across the parallel combination (page 405). This reading will be given by

$$V = I \times R_p = 0.091\,\text{A} \times 13\,\Omega = 1.2\,\text{V} \text{ (to 2 sig. fig.)}$$

The reading of 1.2 V is markedly lower than the value of 2 V calculated before the voltmeter was connected! Connecting a voltmeter across a component in a circuit gives a reading lower than the potential difference existing if the voltmeter were not there. The voltmeter should have a very high resistance, so most of the current passes through the component and very little passes through the voltmeter. In this way, the p.d. across the component stays almost the same (see ITQ5).

A voltmeter should have a very high resistance.

The digital multimeter shown in figure 26.13 can measure current, voltage and resistance. When measuring current, the meter resistance is small (a few ohms or less). When measuring voltage, the meter resistance is extremely high (a few million ohms).

ITQ5

The 20 Ω voltmeter shown in figure 26.12 is replaced by a 20 000 Ω (20 kΩ) voltmeter. Calculate, for this new arrangement: (i) the resistance of the parallel combination; (ii) the circuit current; (iii) the reading on the voltmeter.

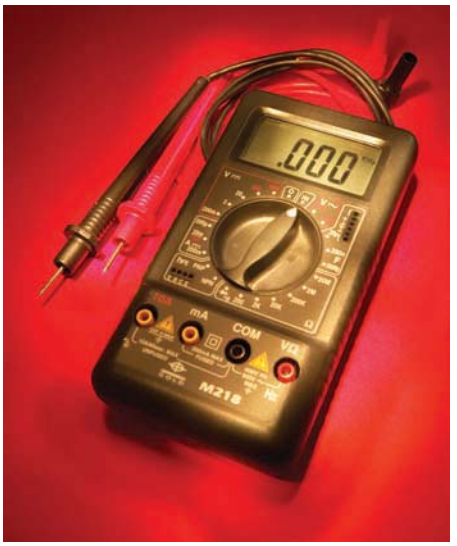


Figure 26.13 A digital multimeter.

Determining current–voltage relationships

Figure 26.14 (a) shows a practical circuit for investigating how current changes according to the voltage applied, for various devices. The variable resistor (rheostat) is wired as a potential divider to provide a variable voltage across the component (figure 26.14 (b)). (Another name for potential divider is 'voltage divider'.)

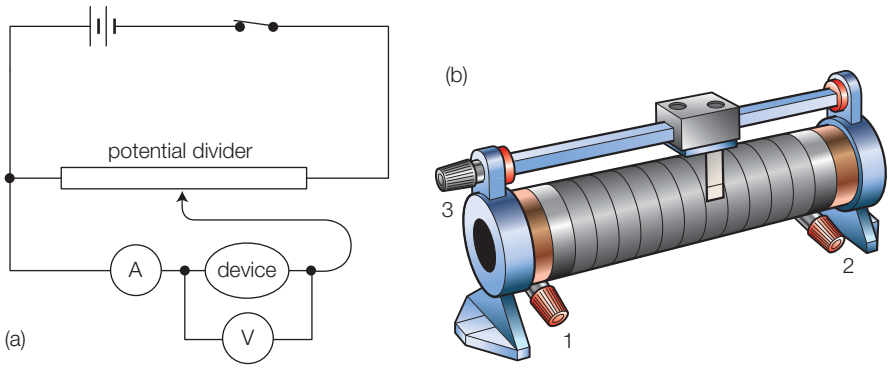


Figure 26.14 (a) Circuit for determining current–voltage relationships for various circuit components. (b) Potential divider showing connections 1 and 2 to battery voltage and output variable voltage connections 1 and 3.

current–voltage characteristic ›

How the current varies with applied voltage can be shown on a graph for each different type of device. The graph is called the **current–voltage characteristic** for that device. Some components show a linear relationship (as evidenced by a straight-line graph) between the current in the device and the voltage across the device. This means that equal changes in voltage result in equal changes in current. Other devices show a non-linear relationship. In the activities below you will be determining the current–voltage characteristics of various devices.

MATHEMATICS:
relations and graphs

Practical activity 26.4

Caution: To be done only under the direct supervision of your teacher.

Current–voltage characteristics of various devices

The following activities are best planned and carried out as a team.

For each of the activities (a)–(d) below, you are to present to your teacher:

(i) a diagram, using standard circuit symbols, of an arrangement suitable for obtaining the current–voltage characteristic for the device (Hint, a starting point might be figure 26.14);

(ii) a list of circuit components needed;

(iii) a list of safety precautions to be taken;

(iv) a list of precautions to be taken to obtain as accurate results as possible;

(v) a list of what will be measured and how many measurements are to be taken;

(vi) an indication as to how variables, if applicable, will be controlled;

(vii) a sketch of the axes of the graph you plan to plot, including units.

After your teacher has approved your plan, carry out the investigations (a)–(d) below under the watchful eye of your teacher.

After you have carried out your experiments and plotted your graphs, describe each graph using terms such as: linear, non-linear increasing, non-linear decreasing. Comment on what the shape of each graph tells you about the resistance of the device. Finally, say whether the device is 'Ohmic', i.e. does it obey Ohm's Law?

(a) *A solid conductor at constant temperature.* A fixed 'resistor' can be used here. Since electric current would tend to heat up a material (see page 394), one of the problems you will encounter is how to keep the conductor at constant temperature. Note: no liquids should be used to accomplish this. (This investigation is actually an Ohm's law investigation, since constant temperature is involved, see page 392.) What does the shape of the graph you obtained for the fixed resistor tell us about the way the *resistance* of the resistor varies with voltage?

reverse voltage ›

High-intensity filament lamps

If a large voltage is applied to a lamp when it is switched on, there is a very large current momentarily, since the cold filament would have a low resistance. The heat produced by this large current could destroy the filament. For this reason, some high-intensity filament lamps (e.g. projector lamps) have a low-voltage 'pre-heat' setting. This voltage allows the filament to heat up, causing the resistance to rise to a certain value, at which point the high voltage is switched on. This 'pre-heat' setting therefore extends the life of the lamp.

forward voltage ›

CHEMISTRY: electrochemistry (ionic solution)

(b) *A solid conductor whose temperature is not constant.* A low-voltage filament lamp could be used here since we know that the temperature of the lamp increases with current. What does the shape of the graph you obtained for the filament lamp tell us about the way the resistance of the filament varies with temperature?

(c) *Semiconductor diode.* The semiconductor diode conducts current readily when a forward voltage is applied. To obtain a **forward voltage** connection, the negative of the voltage supply must be connected to the bar end of the diode (figure 26.15).

Since forward currents as small as 30 mA can cause some semiconductor diodes to burn out, a protective resistor must be placed in series with the diode to limit the current. Worked example 25.3 (page 393) shows how to calculate the value of such a protective resistor in the case of a light-emitting diode. What does the shape of the graph you obtained for the semiconductor diode tell us about the way the *resistance* of the diode varies with forward voltage?

When a **reverse voltage**, increasing from zero, is applied to the diode, there is hardly any current at first. Therefore you are to investigate only the forward voltage case in this activity. (Semiconductors, including the semiconductor diode, are considered in further detail in chapter 30.)

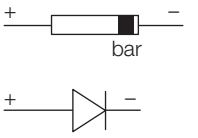
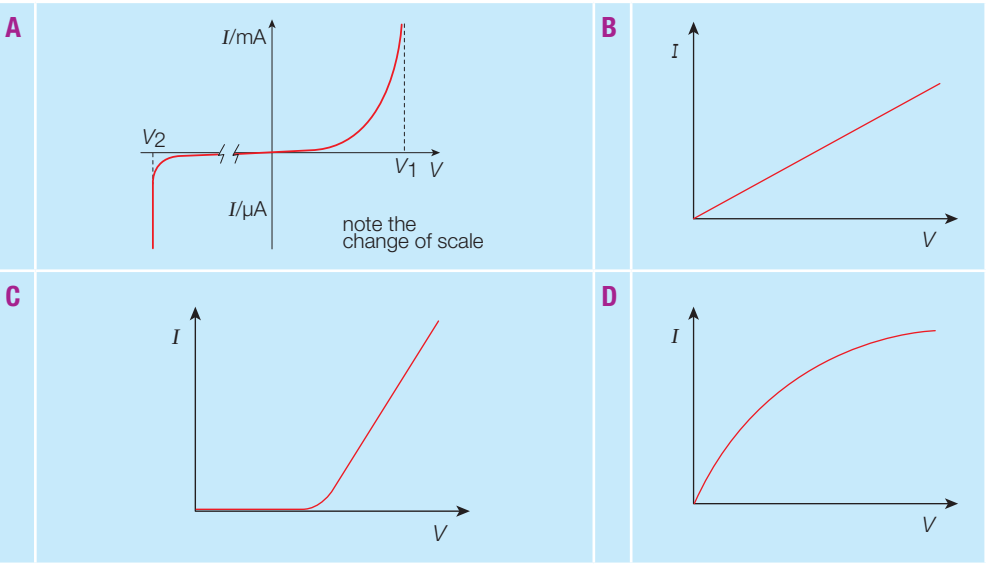


Figure 26.15 Semiconductor diode (the bar end is shown on the right).

(d) *Ionic solution.* In an 'ionic' solution, the charge carriers are ions, that is, atoms that have gained or lost electrons. These are quite unlike the charge carriers in a resistor or a filament lamp (electrons), or in a semiconductor diode (electrons and holes). A suitable investigation can be found in the School-Based Assessments activities near the end of the book (see SBA activity 26.1). What does the shape of the graph you obtained for the ionic solution tell us about the way the resistance of the ionic solution varies with voltage?

ITQ6

The following are graphs of current vs. voltage for the following items: diode, lamp filament, ionic solution resistor at constant temperature. Which graph corresponds to which item? Briefly describe each current–voltage relationship.



Worked example 26.6: The high-voltage overhead power cable

A bare, 100 000 V (10^5 V, or 100 kV) overhead power cable is carrying a current of 1000 A. The resistance per unit length of the cable is 2.0×10^{-5} ohm per meter.

(i) A bird, whose feet are 5.0 cm (5.0×10^{-2} m) apart, is perched on the cable. Estimate the potential difference between the bird's feet.

(ii) A kite, with a string 10.0 m long, is stuck on the bare wire. The resistance of the string is $10^5\,\Omega\,m^{-1} \times 10\,m = 10^6\,\Omega$. Assuming a body resistance of 1000 Ω (i.e. the total resistance of the string and body $\approx 10^6\,\Omega$), estimate the current passing through a person who touches the string and is standing with bare feet on the ground.

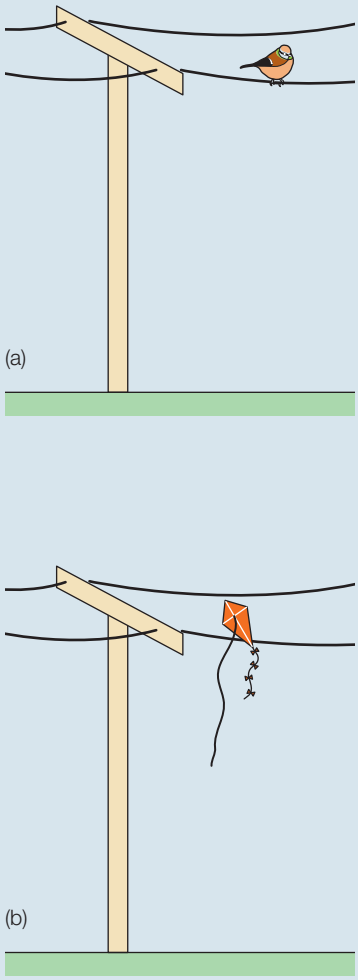


Figure 26.16 (a) The bird is perched on the high-voltage wire and is unharmed. (b) A kite is stuck in the same wire, but touching the string could prove to be fatal.

Solution

(i) The resistance of the wire between the bird's feet is given by

$$(2.0 \times 10^{-5})\,\Omega\,m^{-1} \times (5.0 \times 10^{-2})\,m = 10 \times 10^{-7}\,\Omega$$

The p.d. between the bird's feet is given by

$$\text{p.d.} = \text{current in wire} \times \text{resistance of wire}$$
$$= 1000\,A \times (10 \times 10^{-7})\,\Omega$$
$$= 10^{-3}\,V \text{ (or 1 millivolt, 1 mV)}$$

This voltage is too small to harm the bird, since, applying $V = IR$ to the bird's resistance, R, results in a current, I, of less than 1 mA through the bird.

(ii) The resistance of the kite string is

$$10^5\,\Omega\,m^{-1} \times 10\,m = 10^6\,\Omega$$

Since

$$\text{p.d.} = \text{current} \times \text{resistance}$$

then

$$\text{current} = \frac{\text{p.d.}}{\text{resistance}}$$
$$= \frac{10^5\,V}{10^6\,\Omega}$$
$$= 10^{-1}\,A \text{ (or 100 mA)}$$

This current exceeds 10 mA and could therefore kill a person!

Kites should therefore not be flown near to live overhead electricity cables. The electricity power company should be contacted to deal with situations like this involving overhead power cables.

Energy and power in an electric circuit

In chapter 25, we saw that the energy, E, converted when a charge, Q, flows through a resistor with a potential difference, V, is given by the equation $E = QV$ (pages 390–391). Since current $I = \frac{Q}{t}$, and therefore $Q = It$, and because $V = IR$, we can rewrite that equation in two ways, as follows:

$E = ItV$ joules (J)
$E = I^2Rt$ joules

Power, P, is defined as the rate of using or converting energy. Hence,

$P = \frac{E}{t}$ joules per second (J s^{-1}) (or watts (W))

$P = \frac{ItV}{t}$

We can again write this in two ways:

$P = IV$
$P = I^2R$

Note that energy is measured in joules (J) and power in joules per second (or watts (W)).

Worked example 26.7: Electric stove element

A 500 W element of a 240 V stove is switched on for 1 hour.

(i) How much energy does the element convert in 1 hour?

(ii) What is the current in the element?

Solution

(i) Power, $P = 500\,\text{W}$

Time, $t = 1\,\text{h} = 3600\,\text{s}$

Energy, $E = ?$

Power, $P = \frac{E}{t}$, so energy, $E = Pt$

$E = 500\,\text{W} \times 3600\,\text{s} = 1\,800\,000\,\text{J}$ (or 1.8 MJ)

(ii) Potential difference, $V = 240\,\text{V}$

Since $P = IV$, then $I = PIV$

$I = \frac{P}{V} = \frac{500\,\text{W}}{240\,\text{V}} = 2.08\,\text{A}$

ITQ7

(i) How much energy does a 60 W bulb convert in 1 hour?

(ii) How much current flows through the 60 W bulb when it is operated from a 120 V source?

Chapter summary

- In a series circuit:
  - the current is the same throughout the circuit;
  - the total potential difference, V_s, across a series combination is given by $V_s = V_1 + V_2 + V_3 + \ldots$;
  - the total resistance, R_s, of resistors connected in series is given by $R_s = R_1 + R_2 + R_3 + \ldots$.
- In a parallel circuit:
  - the current entering a junction is equal to the current leaving the junction, i.e. $I = I_1 + I_2 + I_3 + \ldots$;

  - or, the algebraic sum of the currents at a junction is zero, i.e. $I - (I_1 + I_2 + I_3 + ...) = 0$;
  - the potential difference, V_p, across a parallel combination is the same as the p.d. across each component in the combination, i.e. $V_p = V_1 = V_2 + V_3 = ...$;
  - the effective resistance, R_p, of a parallel combination of resistors is given by $\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + ...$.
- When attempting to measure currents and voltages, one should use ammeters of low resistance and voltmeters of high resistance, to avoid changing the values of the currents in, or voltages across, a circuit component.
- A current–voltage graph for a component is called a characteristic for that component.
- Current–voltage characteristics for different components vary: some are linear, some are not.
- We calculate energy, E, and power, P, in electrical circuits using the equations: $E = IVt$ or $E = I^2Rt$ or $P = IV$ or $P = I^2R$

Answers to ITQs

ITQ1 B, switch; C, lamp; D, variable resistor; E, battery; F, fixed resistor.

ITQ2 The rate of flow must be the same or else water would be piling up somewhere in the water circuit.

ITQ3 **(i)** 6 V, since the voltages across all bulbs in series must add up to 120 V.

(ii) The total resistance, $R = 20 \times 5\,\Omega = 100\,\Omega$. Therefore

$$I = \frac{V}{R} = \frac{120\,\text{V}}{100\,\Omega} = 1.2\,\text{A}$$

ITQ4 Since $I = 1.4\text{A}$, and $I = I_1 + I_2$, then $I_2 = 1.4 - 0.8 = 0.6\,\text{A}$

ITQ5 **(i)** $\frac{1}{R_p} = \frac{1}{20\,000} + \frac{1}{40} = 0.02505\,\Omega^{-1}$

$$R_p = 39.9\,\Omega$$

(ii) $I = \frac{3.0}{39.9 + 20} \approx 0.050\,\text{A}$

(iii) $V = IR = 0.050\,\text{A} \times 39.9\,\Omega = 1.995\,\text{V} \approx 2.0\,\text{V}$

Therefore a high-resistance voltmeter gives a more accurate reading of the voltage drop across a low-resistance resistor (when the voltmeter is not present) than does a low-resistance voltmeter.

ITQ6 **(A)** The diode has a low resistance to the forward voltage (steep part of graph) and a very large resistance to the reverse voltage up to voltage, V_2. The increasing slope shows forward resistance decreases with voltage.

(B) Resistor at constant temperature. The graph is a straight line (slope is constant), showing that equal changes in voltage give rise to equal changes in current.

(C) Ionic solution. The solution has a very high resistance until a certain voltage is reached. Beyond this voltage, there is a current which varies linearly with voltage, indicating a constant resistance (a straight-line graph).

(D) Lamp filament. At high voltage, the resistance is large, since the graph is not steep. At low voltages, the steepness of the curve suggests that the resistance is smaller.

ITQ7 **(i)** Power, $P = 60\,\text{W}$

Time, $t = 1\text{ h} = 3600\,\text{s}$

$$\text{power} = \frac{\text{energy}}{\text{time}}$$

$$\text{energy} = \text{power} \times \text{time}$$

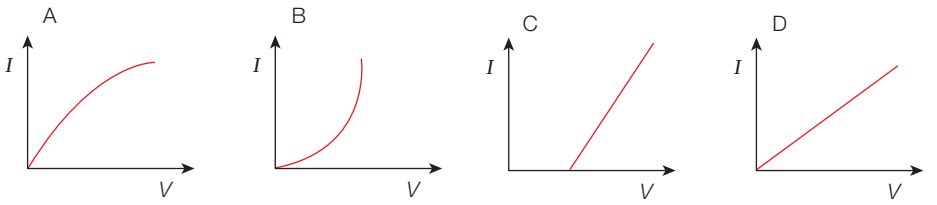
$$= 60\,\text{W} \times 3600\,\text{s}$$

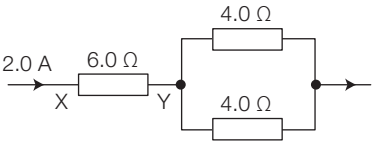
$$= 216\,000\,\text{J} \approx 220\,000\,\text{J (to 2 sig. fig.)}$$

(ii) Power, $P = 60\,W$
Voltage, $V = 120\,V$
As $P = IV$, then
$I = \frac{P}{V} = \frac{60\,W}{120\,V} = 0.50\,A$

Examination-style questions

1 Which of the graphs shows the variation of current, *I*, with applied voltage, *V*, for a semiconductor diode?





The diagram refers to items 2–5

2 The current in *one* of the 4 Ω resistors is:

A 1.0A C 4.0A
B 2.0A D 8.0A

3 The potential difference between X and Y is:

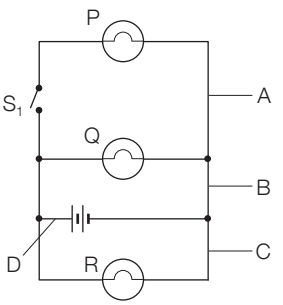
A 8.0V C 16V
B 12V D 20V

4 The power in the 6.0 Ω resistor is:

A 3.0W C 16W
B 12W D 24W

5 The effective resistance of the circuit is:

A 6.5 Ω C 14 Ω
B 8.0 Ω D 22 Ω



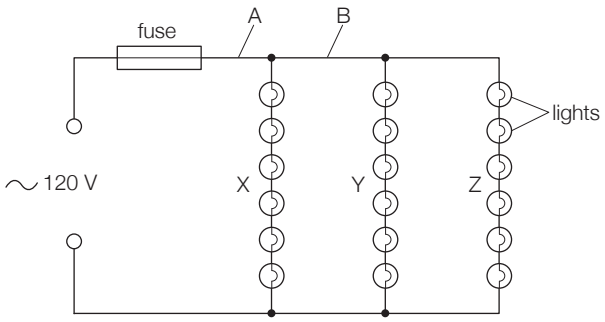
The diagram refers to items 6 and 7

6 Which of the lamps will be lit up after S_1 is closed?

A P only
B P and Q only
C P, Q and R
D Q and R only

7 With S_1 closed, where should another switch S_2 be placed so that when S_2 is open only R lights?

8 The diagram shows three identical strings of lights. Each lamp requires 0.50 A to operate properly.



(i) (a) How are the lights connected within each string?

(b) How are the strings of lights connected to the 120 V supply?

(ii) If the lamps are operated correctly, determine the current:

(a) in string X;

(b) along the wire at A;

(c) along the wire at B.

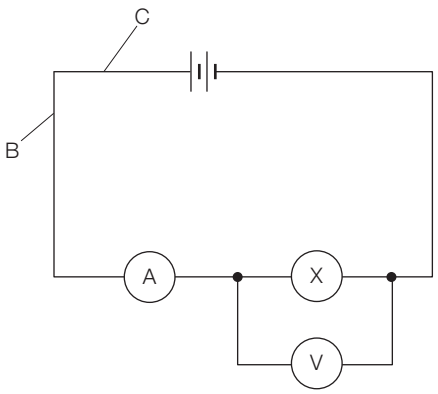
(iii) Calculate, for string X only:

(a) the total resistance of the lamps;

(b) the total power of the lamps.

(iv) Determine the rating of a suitable fuse for the circuit shown. Briefly discuss reasons for the particular rating of fuse chosen.

(v) When all the lamps are lighting, a lamp in string Y blows and forms an open circuit. Describe what would be observed in strings X, Y and Z.



9 The diagram shows a circuit for measuring current through, and voltage across, a device X.

(i) What component should be inserted in the circuit

(a) at C, for safety reasons?

(b) at B, to vary the current in the circuit?

(ii) The readings in the table were obtained in an experiment, using the above circuit, with the devices at B and C inserted and the circuit completed.

Current, *I* (mA)	5	11	21	38	50	75	120
Voltage, *V* (V)	0.2	0.4	0.6	0.8	0.9	1.0	1.1

(a) Plot a graph of current, *I*, on the vertical axis against voltage, *V*, on the horizontal axis.

(b) Based on the shape of your graph, suggest what component X might be.

(c) Suggest a safe maximum voltage range over which X can be operated.

10 A series circuit consists of a battery, B, of negligible internal resistance, a switch, S, a resistor, R, and an ammeter, A. A voltmeter, V, is connected across R.

(i) Using circuit symbols, draw the circuit described above.

(ii) A resistor, R, identical to R, is inserted in series with the battery. What would happen to the reading on the ammeter? (Assume that the ammeter resistance is very low and the voltmeter resistance is very high.)

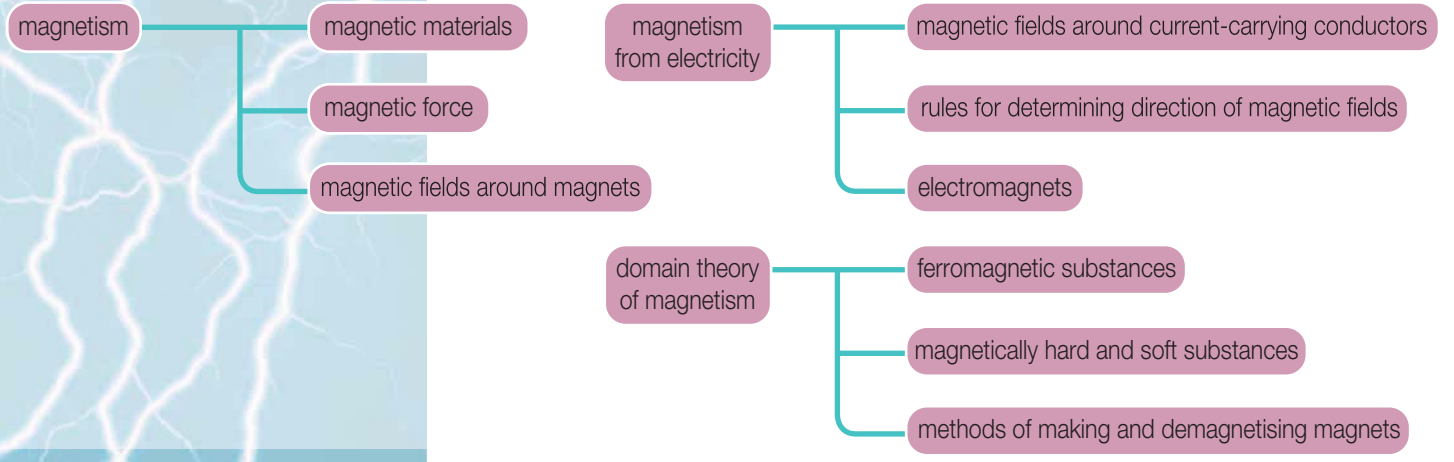
27 Magnetism and Electricity

By the end of this chapter you should be able to:

- differentiate between magnetic and non-magnetic materials
- describe a simple experiment to identify the poles of a magnetic dipole
- describe simple experiments that show that there is a repulsive force between like poles and an attractive force between unlike poles
- explain how a magnet can attract an unmagnetised object
- distinguish between materials used for making permanent and temporary magnets
- investigate the effects of the separation of magnets and polarity on the magnitude of the force between them
- define a magnetic field as the region around a magnet in which a magnetic force may be exerted
- recall that a magnetic field line indicates the direction of the magnetic force acting on a north-seeking pole (N-pole)
- map the magnetic field (i) around a single strong magnet and (ii) formed by two strong magnets
- draw simple diagrams to show how permanent magnets can be used to create a uniform magnetic field over a small region
- describe simple experiments to investigate the magnetic field around current-carrying conductors
- apply rules that relate the direction of current flow to the direction of the associated magnetic field
- sketch the magnetic field patterns around current-carrying conductors;
- construct a simple electromagnet
- describe a commercial application of an electromagnet
- explain the action of a simple magnetic relay
- account for magnetism using the domain theory
- describe methods of magnetisation and demagnetisation



Figure 27.1 This powerful electromagnet, hoisting heavy scrap metal, is an application of the close association between current electricity and magnetism.



The close association between magnetism and electricity

magnetite magnet ›

More than 2000 years ago, the ancient Greeks discovered a type of rock that had a property of attracting iron. The rock was discovered in Magnesia and so it was called **magnetite**, from which we have the word '**magnet**'. One thousand years later, Chinese sailors found that, when this type of rock shape was elongated, and was suspended, it settled in an approximately north–south direction. They used this property to make a simple compass for their sea voyages. The rock became known as **lodestone**, which means 'the leading stone'.

Internet search term: history of magnets

lodestone ›

pole ›

At first, it was thought that magnetic properties were caused by discrete entities, called north and south '**poles**', within materials. The 'poles' were thought to behave similarly to the positive and negative charges present in matter. Studies of magnetism, however, have shown that separate magnetic north and south poles do not exist. Other studies have shown that magnetism also results from the flow of charge – that is, from electric currents.

In this chapter we first look at some basic concepts in magnetism, and then at some introductory examples of the close association between magnetism and current electricity. The electromagnetic hoist shown in figure 27.1 is one application of this relationship.

Magnetism

Magnetic materials

magnetic material ›

A **magnetic material** is one that is capable of being magnetised. Metals such as iron, nickel and cobalt are magnetic materials. When a magnetic material is placed near to a magnet, it will itself become magnetised and therefore be attracted to the magnet.

A magnetic material is one that is capable of being magnetised.

A magnetised object will attract another magnetic material such as iron. So, to find out whether an object is magnetised, we can simply dip it into a heap of iron filings. If filings stick to the object when it is removed from the heap, we infer that the object is magnetised (figure 27.2). We call a magnetised object a magnet.

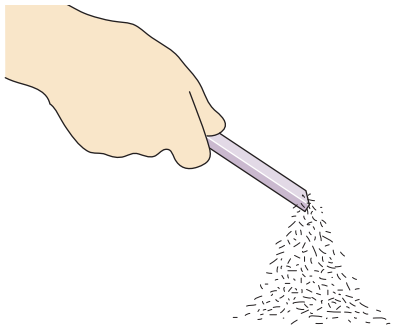


Figure 27.2 Testing an object to find out whether it is magnetised.

Practical activity 27.1



Magnetised objects and magnetic materials

1 Discuss the difference between a *magnetic* material and a *magnetised* object.

2 Plan and carry out activities to determine whether metals (such as copper, aluminium, iron, zinc, nickel) are magnetic materials, and whether objects made from these materials (such as nails, paper clips, foil, coins), are magnetised.

3 Repeat item 2 above by investigating non-metals (such as such as wood, plastic, rubber and glass).

Question

Were there any objects that were not magnetised but yet were made of magnetic material? How did you know this?

Magnetic poles

Bear in mind that although we refer to 'poles' as being near the ends of magnets, separate magnetic 'poles' do not exist. The so-called 'poles' are formed from the alignment of tiny atomic 'dipoles' (see domain theory, page 425). The breaking of magnets into two (figure 27.3) shows that we cannot separate an N-pole from an S-pole.

north-seeking pole or N-pole ›
south-seeking pole or S-pole ›
dipole ›

A pole is that part of a magnet where the magnetism appears to be concentrated. If a bar magnet is dipped into iron filings, most of the filings are attracted towards the ends of the magnet (see figure 27.3). This shows that the poles of the bar magnet are situated near its ends.

When suspended horizontally near the Earth's surface, a bar magnet settles in an approximately north–south direction. The end of the magnet that points north is called the **north-seeking pole**, or **N-pole**; and the end pointing south is the **south-seeking pole**, or **S-pole**. Because the bar magnet has two poles, it is sometimes referred to as a magnetic **dipole**.

If the bar magnet is broken into two pieces, as in figure 27.3, iron filings are attracted towards both ends of each piece! Like the original magnet, each piece has two poles. When the pieces are suspended horizontally, each settles in a north–south direction, so each piece has an N-pole and an S-pole. In other words, each piece is a magnetic dipole. It is thought that magnets exist only as magnetic dipoles. Magnetic monopoles are not believed to exist.

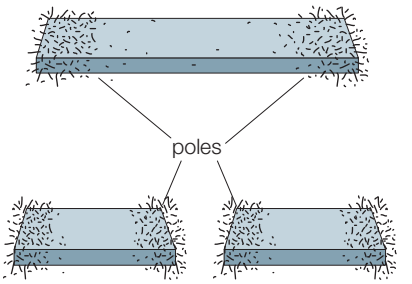


Figure 27.3 Magnetic poles.

Magnetic forces

Practical activity 27.2

Investigating magnetic forces

1 You are given two bar magnets with their poles marked. First, carry out an activity to confirm that the poles are marked correctly.

2 Using the arrangement shown in figure 27.4 (or just placing the magnets on the desk surface), observe and record the directions of the magnetic forces when:

(a) a N-pole is brought near to another N-pole;

(b) a S-pole is brought near to a S-pole;

(c) a S-pole is brought near to a N-pole.

3 Investigate the relationship between the magnetic force and the distance between two magnets.

MATHEMATICS: relations and functions

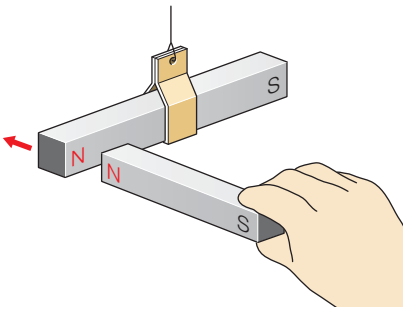


Figure 27.4 Investigating magnetic forces.

Experiments have shown that when magnets exert forces on each other:

- Like poles repel; unlike poles attract
- The force decreases as the distance between the magnets increases, and vice-versa. This is called a kind of inverse relationship.

Magnetic fields around magnets

magnetic field ›

A **magnetic field** is a region in which a magnetic force acts.

plotting compass ›

Magnetic fields can be investigated using a **plotting compass**, which is a small, very light magnet suspended on a low-friction pivot. The magnet of this compass (called the compass needle) always aligns itself in the direction of the field in which it is placed. As with all magnets, the end of the needle that points north when freely suspended is a north pole.

ITQ1

How does the magnetic force between two magnets, X and Y, change as the distance between them decreases?

Confirming the identity of a pole

We do not use attraction to identify poles, since, as we shall see, a given pole of a magnet can attract both ends of some types of objects, by induced magnetism.

Since like poles repel, we can use a repulsion test to identify a magnetic pole. If the N-pole of a magnet repeals the end, P, of another magnet, then end P is also an N-pole. If the end, P, repels the N-pole of a magnetic compass needle, the end, P, is therefore a N-pole.

To investigate the pattern of a magnetic field, the compass is placed in the field as shown in figure 27.5, and two dots are marked on the paper next to the N-pole and S-pole of the magnetic compass needle (position 1 in figure 27.5). The compass is then moved to position 2 and another dot is made at the unmarked end, and so on. When the dots are joined sequentially, a pattern showing **magnetic lines of force** in the magnetic field is obtained.

magnetic lines of force ›

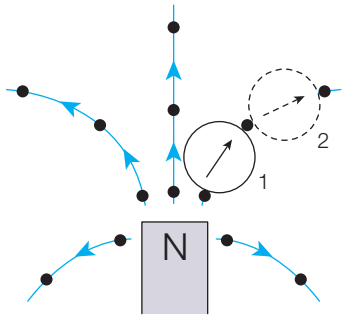


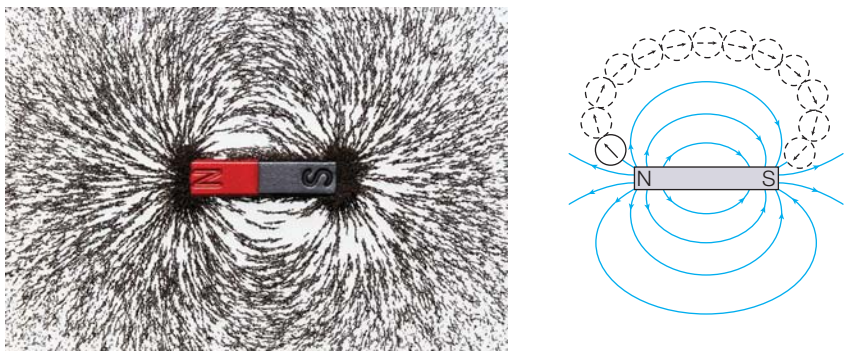
Figure 27.5 Plotting a magnetic field. The arrows drawn on the lines of force (field lines) show the direction in which an N-pole would move, if it existed.

A magnetic line of force, or field line, shows the path in which a tiny N-pole (if it existed) would move when placed in a magnetic field. The arrows on the lines of force indicate the direction in which an N-pole would move. So the magnetic needle of a plotting compass will align itself along a line of force. A set of lines of force show the magnetic field pattern due to the magnet.

The magnetic fields shown here do not include the effects of the Earth's magnetism.

Figure 27.6 shows the magnetic field patterns obtained around a single bar magnet, using a plotting compass and also iron filings. Experiments show that the magnetic force is strongest where the field lines are closest together. Magnetic field lines crossing an area perpendicularly represent a quantity, which we call **magnetic flux**. Magnetic flux is discussed further in chapter 28.

magnetic flux ›



(a) (b)

Figure 27.6 The magnetic field around a single bar magnet: (a) using iron filings sprinkled close to the magnet; (b) using a plotting compass.

Figure 27.7 shows the field pattern obtained with two bar magnets with north poles facing each other. Note that lines of force do not cross. At point X there is no net magnetic force. X is called a **neutral point** in the magnetic field.

neutral point ›

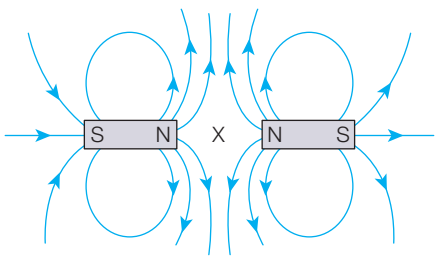


Figure 27.7 The magnetic field due to two bar magnets with N-poles facing each other.

Practical activity 27.3

Mapping magnetic fields

In this activity, you are to predict and make sketches of your prediction for each magnetic field, below. Then, using a plotting compass, along with a pencil and a sheet of paper, map the magnetic fields due to the following, and identify neutral points, if any.

1 *The Earth.* (Make sure you do this activity away from magnetic materials or magnetised objects.) Obtain the horizontal component of field lines due to the Earth's magnetism.
2 *The Earth and a bar magnet.* Place the bar magnet parallel to the horizontal component of the Earth's field with the magnet's N-pole pointing south.
3 *Two S-poles.* Place two S-poles end-to-end about 10–15 cm apart.
4 *An N-pole and an S-pole end-to-end.* Place the two poles end-to-end about 10–15 cm apart.
5 *Two magnets parallel to each other.* Place two bar magnets side by side about 10–15 cm apart with poles oriented in the same direction.
6 *Two magnets anti-parallel to each other.* Place two bar magnets side by side about 10–15 cm apart with poles oriented in opposite directions.

ITQ2

In which of the arrangements, 1–6, were there neutral points?

ITQ3

In the horseshoe magnet shown in figure 27.8, which letter represents a region in which the magnetic field is most nearly uniform?

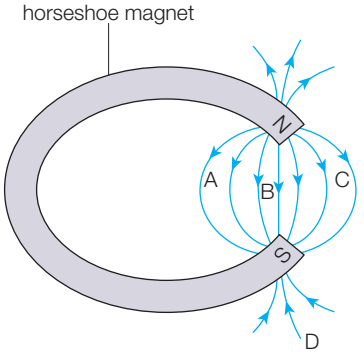


Figure 27.8 Magnetic field between the poles of a horseshoe magnet.

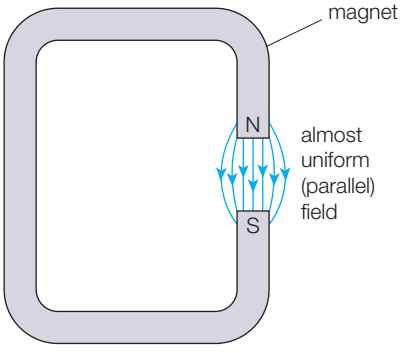


Figure 27.9 A strong, fairly uniform field produced by a C-shaped magnet.

Uniform and non-uniform magnetic fields

A uniform magnetic field is one in which the magnetic field lines are parallel. Figures 27.8, 27.9, 27.11 and 27.12 all show examples of non-uniform fields.

A fairly uniform magnetic field can be obtained in the gap between the N-pole and S-pole of a 'C'-shaped magnet (figures 27.8 and 27.9). When the gap is very small, a nearly uniform field is found within and just outside the gap. Since most of the magnetic field is confined to this small region, it can be very strong. Magnetic tape recording heads and electromagnetic 'read-write' heads on computer disk drives (figure 27.10) make use of such gaps to produce strong magnetic fields that can magnetise tapes or disks as they pass close to the gaps.

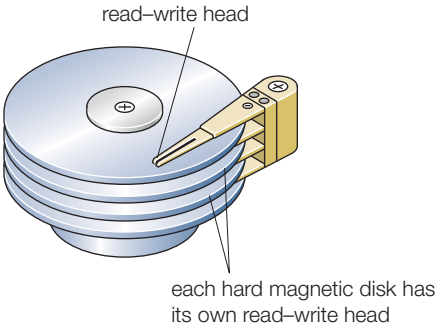


Figure 27.10 Hard disk of a computer showing the 'read-write' heads. The read-write head is moved by the computer to 'write' or record information on an unused area of the disk, or to 'read' or retrieve previously recorded information.

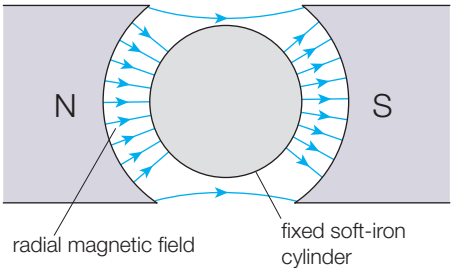


Figure 27.11 The radial magnetic field of a moving-coil galvanometer.

In the moving-coil galvanometer, a cylindrical soft-iron core is placed in the gap between curved N- and S-poles to produce a field that is radial (figure 27.11) within the gaps. The principle of operation of the galvanometer is discussed in chapter 28.

The Earth as a magnet

The fact that a magnet suspended near the Earth always points in a north–south direction shows that the Earth itself is behaving as if it were a magnet! Because the north-seeking or N-pole of a suspended magnet points towards the geographic North Pole, the 'S-pole' of the Earth's magnetic field must be near the geographic North Pole, as shown in figure 27.12.

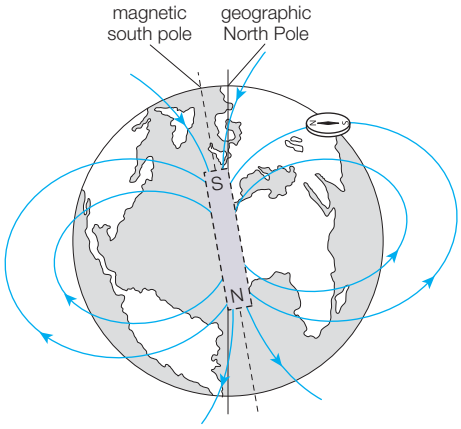


Figure 27.12 The Earth's magnetic field. (Although the Earth acts as though there were an enormous bar magnet inside it, such a magnet could not exist – the Earth's core is far too hot!)

The earth's magnetic field is of great importance since charged particles given off by the Sun are trapped in this field and are hindered from causing damage to life on Earth. Thus, a shield from high-speed, charged particles exists around the Earth. Migrating birds have biological magnetic sensors and use the Earth's magnetic field for navigation (figure 27.13).

Internet search term: navigation of migrating birds

BIOLOGY: species survival adaptations

The magnetic poles of the Earth are slowly shifting. In a few hundred years, the Earth's magnetic S-pole, currently west of the geographic North Pole, is expected to be east of the geographic North Pole.

Internet search term: Earth's shifting magnetic poles



Figure 27.13 Migrating birds navigate using the Earth's magnetic field.

Magnetism from electricity

Oersted's discovery

In 1819, the Danish scientist, Hans Oersted, set up an experiment to show that magnetism had nothing to do with electricity. Surprisingly, the experiment proved the opposite – current electricity and magnetism were inseparable! Whenever an electric current flowed along a wire, there was a magnetic field around the wire.

Figure 27.14 shows an apparatus that demonstrates this effect. Each time the current is switched on, the compass needle is deflected, indicating the presence of a magnetic field (see ITQ4).

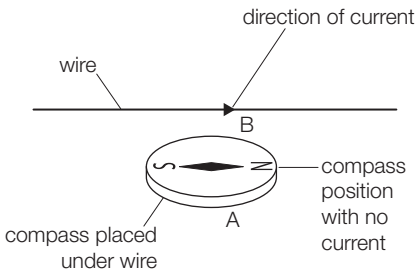


Figure 27.14 An apparatus for demonstrating Oersted's discovery.

Direction of the magnetic field around a wire carrying a current

right-hand grip rule ›

Experiments have shown that the direction of the magnetic field due to a current-carrying conductor can be predicted by applying the **right-hand grip rule**. Imagine that the wire is gripped with the right hand so that the thumb points in the direction of the (conventional) current; the fingers then curl in the direction of the field, as shown in figure 27.15 (a).

right-hand corkscrew rule ›

The **right-hand corkscrew** rule also gives the direction of the magnetic field due to a current. If the forward direction of travel of a corkscrew is the same as the direction of the current, then the direction in which the corkscrew is being turned gives the direction of the magnetic field (figure 27.15 (b)).

ITQ4

In which direction, A or B, will the N-pole of the compass in figure 27.14 move when current flows as shown in the wire?

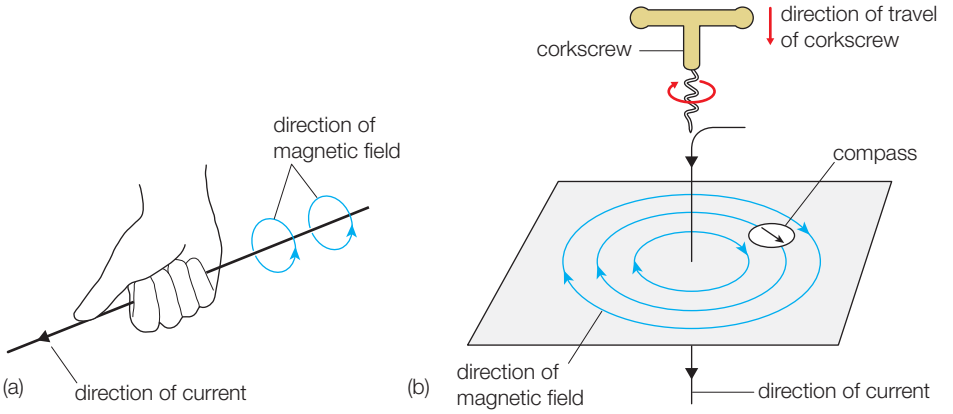


Figure 27.15 (a) The right-hand grip rule. (b) The right-hand corkscrew rule.

In diagrams, we use dots and crosses to represent the directions of fields, or currents at right angles to the page. A dot (•) indicates a direction coming out of the page at right angles. A cross (x or +) indicates a direction into the page at right angles. Dots or crosses *with circles* around them are used for currents, and *without circles* for magnetic fields. Figure 27.16 shows the magnetic fields due to straight current-carrying conductors that are perpendicular to the page, the current directions are represented by a circled dot and a circled cross.

The dots and crosses convention

Think of an arrow coming out of the page, and one going into the page. In the first case, we see the arrow tip (the dot) and not the feathers. In the second case, we see the feathers forming an 'x' or an '+' and not the tip.

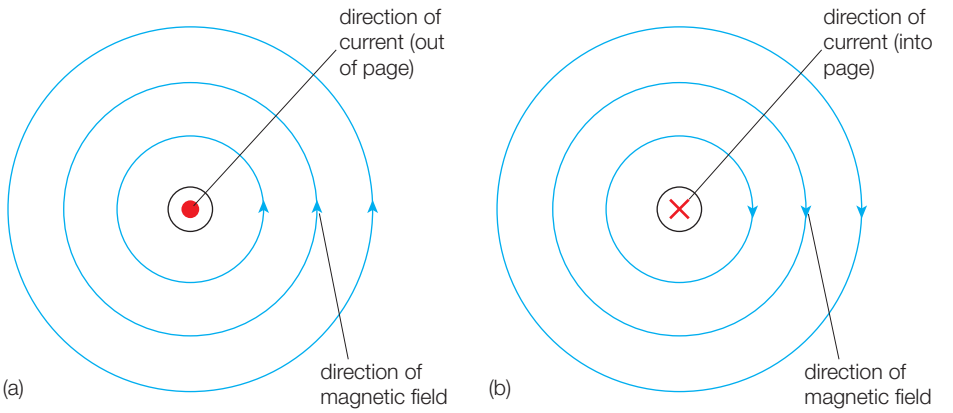


Figure 27.16 Magnetic field around a straight current-carrying conductor. (a) Current direction out of the page. (b) Current direction into the page.

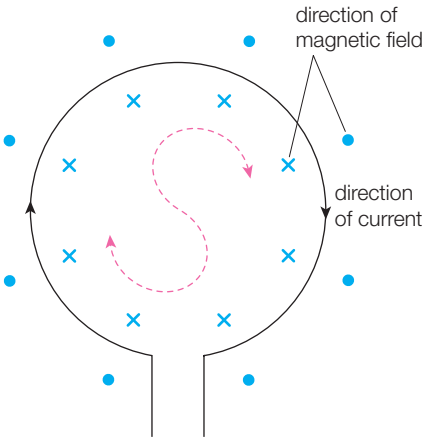


Figure 27.17 Magnetic field due to a clockwise current in a circular wire. The dashed 'S' shows another way of remembering that a clockwise current is associated with an S-pole.

Figure 27.17 shows a circular coil with a *clockwise*-flowing current. Imagine that we grip the coil with the right hand and, applying the right-hand grip rule we find that within the coil the direction of the magnetic field is into the page. Hence, this face of the coil is an S-pole, since magnetic lines of force are directed towards an S-pole. We might remember this using the 'CS' rule: a clockwise current is associated with an S-pole. (Figure 27.17 also shows another way of remembering this: arrows on the ends of an 'S' drawn inside the coil points in the direction of the clockwise current.)

Viewed from the other side (the back) of the same coil in figure 27.17, the current is *anticlockwise*. This side behaves like a N-pole, since the direction of the magnetic field within the coil now points toward the observer. We remember this using the 'AN' rule: an anticlockwise current is associated with an N-pole. (Another way of remembering this is that arrows on the ends of an 'N' drawn inside the coil points in the direction of the anticlockwise current.)

For the same current in a coil, we therefore see that one side of the coil is a N-pole, while the other side is a S-pole. We cannot obtain magnetic monopoles from currents in coils – we obtain only dipoles. This might be a clue as to

why magnetic monopoles are not believed to exist, as we mentioned earlier – perhaps, magnetism is due to circular currents, after all.

ITQ5

Using dots and crosses near A, B and C, show the directions of the magnetic fields due to the straight and circular wires carrying current as shown in figure 27.18.

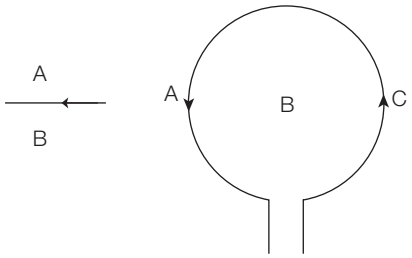


Figure 27.18

ITQ6

In figure 27.19, what pole is the girl facing due to the current shown? What pole is the boy facing?



Figure 27.19 A single loop of wire carrying a current, viewed from opposite sides.

Investigating magnetic fields around current-carrying conductors

Caution: because of the sizes of the currents involved, these investigations should be done by the teacher and not by students.

Figure 27.20 shows an arrangement for investigating the magnetic field pattern due to a current in a straight wire. A stiff piece of cardboard is supported horizontally on a wooden desk (or using wooden clamps and stands), away from magnetic materials. When the current from a low-voltage supply is switched on, a plotting compass is used to determine the direction of the magnetic field. The field pattern is obtained by marking dots as near as possible to the ends of the compass needle and subsequently joining consecutive dots.

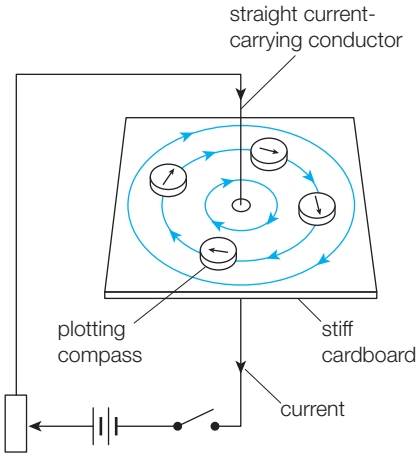


Figure 27.20 Apparatus for investigating the field pattern due to a current in a straight wire.

A typical field pattern, consisting of concentric circles, is shown in figure 27.20. If the direction of the current is reversed, the direction of the magnetic field also reverses, while the circular pattern remains.

The field due to a flat, circular current-carrying coil can be investigated using the apparatus shown in figure 27.21. A flat, circular coil is made by winding insulated wire around a **former** such as a plastic tube or the hollow cardboard cylinder of a tissue paper roll. A safe current, set by the teacher, is passed through the coil. The direction of the field is obtained as before by using a plotting compass. Again, the field changes direction when the current is reversed in the coil.

former

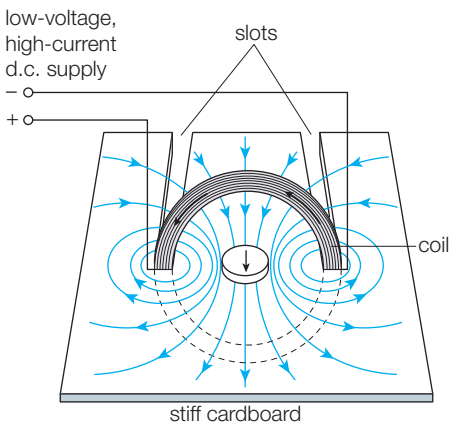


Figure 27.21 Investigating the magnetic field due to a current in a flat, circular coil of wire. (The diameter of the coils is about 5 cm.)

A longer coil or 'solenoid' can be made of stiff copper wire wound as shown in figure 27.22. (A cardboard or plastic tube could be used as the former.) Using the arrangement shown in figure 27.22, the direction of the magnetic field can be plotted on the sheet of stiff cardboard, as described above. A typical magnetic field pattern for a solenoid is shown in the figure.

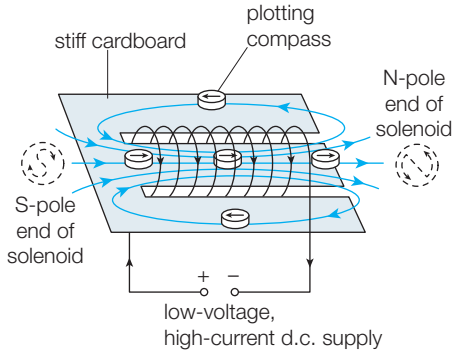


Figure 27.22 Arrangement for investigating the field pattern due to a current in a solenoid.

Electromagnets

We saw earlier that a single loop of wire when viewed from one side shows current flowing clockwise around the loop (figure 27.19). This side therefore behaves like an S-pole (from the 'CS' rule: see page 422). The same loop, when viewed from the other side, shows the same current flowing anticlockwise around the loop! This side behaves like an N-pole (from the 'AN' rule). Thus, whenever an S-pole is formed due to a current, an N-pole is also formed. The current-carrying loop is said to form an **electromagnet**. As we have just seen, such electromagnets appear only as magnetic dipoles, so a coil carrying a direct current behaves like a magnet.

electromagnet ›

A coil of insulated wire connected to a battery is therefore an electromagnet. The end where the current flows in a clockwise direction behaves like an S-pole. The other end, where the current flows in an anticlockwise direction, behaves like an N-pole.

An electromagnet can be made powerful as follows:

1 By using a large number of turns of wire to make up the coil. Each turn gives rise to a magnetic effect in the same direction, and the magnetic effects add to produce a strong field.

2 By placing a soft-iron core inside the coil, as in figure 27.23. (This is because of the realignment of magnetic 'domains' within the soft iron, under the influence of the magnetic field produced by the current in the coil (figure 27.29 – we will discuss this later in this chapter.)

3 By using a very large current in the coil. (The greater the current, the greater the magnetic effect produced.)

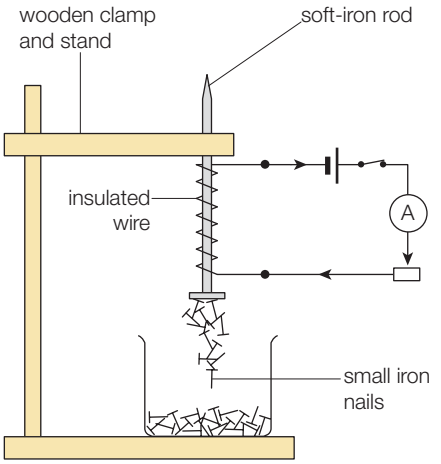


Figure 27.23 An electromagnet (or solenoid) with an iron nail as core. When current flows in the coil, the magnetic field produced is strong enough to lift several small iron nails.

Soft iron is a form of iron that is very easily magnetised and demagnetised. Magnetically 'soft' and 'hard' materials are discussed on the following pages.

Uses of electromagnets

electromagnetic relay ›

One very common use of electromagnetism is the **electromagnetic relay** switch (figure 27.24). When a small current flows in the coil wrapped around the soft-iron core, the coil becomes a powerful electromagnet because it is made of many turns of wire wrapped around a soft-iron core. The electromagnet attracts a pivoted soft-iron armature, which, in turn, pushes the springy contacts together. The terminals of the springy contacts are connected to a separate circuit. The contacts act as a switch to open and close this separate circuit.

Relays are often used to switch on circuits carrying large currents. In a motor car starter circuit (figure 27.25), a small current of about 1 A activates the relay coil at switch-on. The electromagnet produced attracts the iron plunger towards the open relay contacts. The relay contacts are bridged, in turn switching on the starter motor circuit, which draws over 60 A of current.

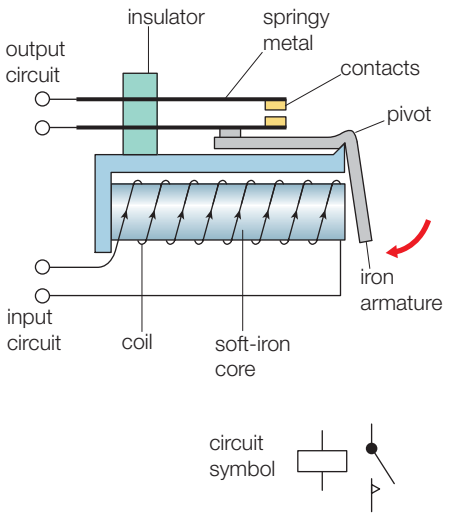


Figure 27.24 Electromagnetic relay switch.

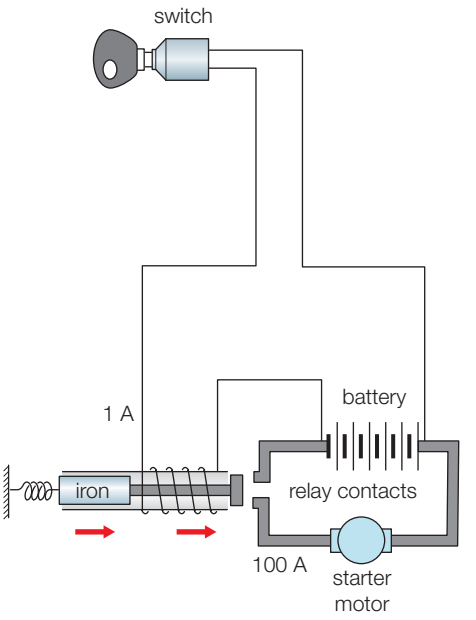


Figure 27.25 Relay switching circuit for a car's starter motor.

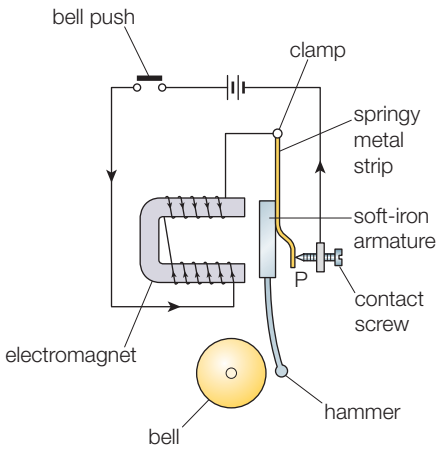


Figure 27.26 An electric bell.

'Relay switching' is used because it is far safer to directly operate a switch controlling 1 A than one controlling 60 A.

Electromagnets are also used in electric bells. When the switch in the bell circuit is closed (figure 27.26), the electromagnet attracts the soft-iron armature, and the hammer strikes the bell. The contact is broken at P, causing the coil to lose its electromagnetism. The spring then pulls the armature back to its original position and makes contact at P, now closing the circuit again. The cycle is automatically repeated, so the armature and hammer move to and fro. The hammer repeatedly hits the bell as long as the switch remains closed.

Very powerful electromagnets are used for lifting scrap iron (figure 27.1, page 416). On a very different scale, eye surgeons sometimes use electromagnets to remove specks of iron from people's eyes after accidents.

Domain theory of magnetism

Ferromagnetic materials

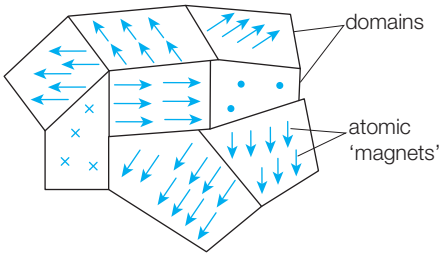


Figure 27.28 Magnetic domains consist of 'atomic dipoles' that are aligned in the same direction.

It is suggested that magnetism is caused by moving charges. As an electron orbits an atomic nucleus, a loop of electric current is produced. This results in the formation of a tiny magnet dipole (represented in figure 27.27 by the N–S arrow). If as many electrons orbit one way as orbit the opposite way, the magnetic dipole effects are opposite and cancel each other out. This explains why most substances appear non-magnetic.

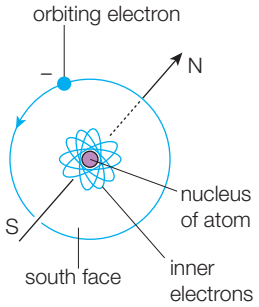


Figure 27.27 Orbiting electrons cause an atom to behave like a tiny magnet with N- and S-poles.

In some substances, like iron, nickel and cobalt, however, groups of atoms have their magnetic dipoles aligned together in 'domains' (figure 27.28). Within each **domain**, the magnetism due to each of the atoms adds up to produce a large magnetic effect. When more domains are aligned in one direction than in another, the material as a whole will behave as a magnet. Substances in which this happens are called **ferromagnetic** materials.

domain ›

ferromagnetic ›

Magnetic domains within an object made of a ferromagnetic material can be enlarged by cooling the object. As the object cools, less energy is available for atoms to vibrate randomly. Thus, more atomic dipoles line up in the same directions, forming larger domains. The magnet becomes more powerful. Heating the object causes the particles to vibrate faster. The domains then become smaller because atomic dipoles take up different directions on account of the increased vibrations.

Magnetic domains in a ferromagnetic substance can also be enlarged by placing the substance in a magnetic field. Under the influence of the field, many atomic dipoles line up in one direction. Domains merge to produce larger ones.

IT Q7

On the basis of the domain theory, would you expect a ferromagnetic substance to remain magnetic at very high temperatures? Explain your answer.

IT Q8

Certain substances show only a slight degree of magnetism at room temperature. How can we explain this?

'Hard' and 'soft' magnetic materials

magnetically hard material ›

permanent magnet ›

A material is said to be **magnetically 'hard'** if it is difficult to change its domain structure. Such substances are hard to magnetise and they are also difficult to demagnetise. Because they tend to retain whatever magnetism they have, they are often used in making **permanent magnets**. Powerful permanent magnets are usually made from alloys such as steel (made from carbon and iron), alnico (made from aluminium, nickel and cobalt), and magnadur (made from oxides of iron and barium). Permanent magnets are therefore made from magnetically hard materials.

Permanent magnets are used in loudspeakers and in small electric motors (see chapter 28). Permanent magnets are also placed on the doors of refrigerators to keep the doors shut tightly against the iron body of the fridge.

magnetically soft material ›

Magnetically 'soft' materials are magnetised very readily, and also lose their magnetism very easily. If soft iron is placed in a magnetic field, the domains readily line up in the direction of the field (figure 27.29). The soft iron becomes a strong magnet since most of the domains are aligned in the direction of the magnetising field. We call this 'induced magnetism'. When the magnet field is removed, the thermal energy within the iron is enough to cause the 'atomic dipoles' to vibrate so that the domains are oriented once more in random directions. The soft iron therefore loses the net magnetism it had gained. Materials like soft iron are used for making **'temporary' magnets** or electromagnets.

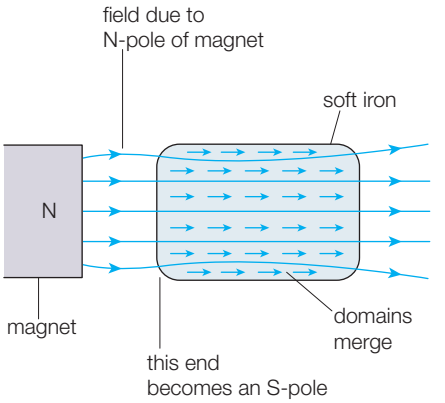


Figure 27.29 Magnetic domains within soft iron align readily with the field from a bar magnet. The soft iron therefore becomes a magnet and is attracted to the bar magnet.

Internet search terms: hard and soft magnetic materials

temporary magnet ›

Temporary magnets are used extensively with magnetising fields created in electrical circuits, for example in relays and electric bells. They are also used in magnetic 'read-write' heads used for transferring information in computers (see figure 27.10). Specialized use is made of the magnetically soft *mumetal* (made from nickel and iron) in magnetic shielding of circuits that are affected by low-frequency electromagnetic fields.

ITQ9

The N-pole of a permanent magnet attracts the end, A, of a metal bar. What type of magnet might the metal bar be?

Methods of making magnets

Stroking

If a piece of steel is stroked repeatedly by a magnet, as shown in figure 27.30, the steel becomes magnetised. The magnetic domains within the steel become aligned, thus producing a strong magnet. The domains remain in place because steel is magnetically hard.

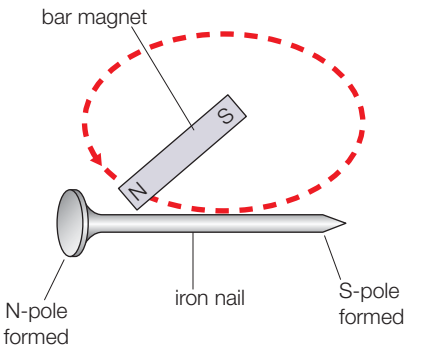


Figure 27.30 The stroking method of making a magnet.

Electrical method

If a piece of steel is placed inside a coil of many turns carrying a large direct current, the steel becomes magnetised (figure 27.31). The domains within the steel become aligned and remain so because steel is magnetically hard. Today, permanent magnets are generally made by the electrical method.

A piece of soft iron, placed inside a coil carrying a direct current, easily becomes magnetic as its domains realign. As mentioned earlier, as soon as the field is removed (i.e. the current is switched off), the soft iron loses most of its magnetism. Temporary magnets made by the electrical method are called solenoids or electromagnets.

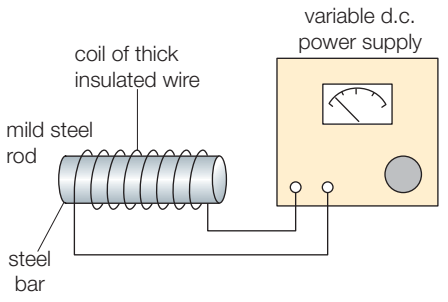


Figure 27.31 Electrical method of making a permanent magnet.

Demagnetising a magnet

Demagnetising a magnet involves randomly disrupting the alignment of the domains. This can be done by repeated mechanical shock, for example by dropping the magnet onto a hard surface. Heating the magnet and allowing it to cool in an east-west direction leads to demagnetisation, since thermal vibrations lead to the domains becoming randomly oriented. A magnet can be demagnetised by placing it inside a solenoid carrying alternating current, with its axis in an east-west direction, and pulling it slowly out of the coil. The alternating current produces an alternating magnetic field which causes the ordered domains in the magnet to become randomly rearranged, resulting in a net magnetisation of the material that is practically zero.

Chapter summary

- A magnet is an object that can attract iron.
- A magnetic pole is that area of a magnet where the magnetism may be considered to be concentrated. Magnetic poles occur in pairs, called dipoles.
- When a magnet is freely suspended, the end that points roughly towards the geographic North Pole is called the N-pole or north-seeking pole. The end pointing towards the geographic South Pole is called the S-pole or south-seeking pole.
- Experiments with magnetic poles show that like poles repel and unlike poles attract.
- The region in which a magnetic force acts is called a magnetic field. The direction of a magnetic field is the direction in which a tiny N-pole (if it existed) would move if placed in the field.
- Magnetic forces are strongest where field lines are closest.
- A magnetic field is produced when electric current flows.
- The direction of the magnetic field produced by a current in a straight or circular conductor can be determined using the right-hand grip rule or the right-hand corkscrew rule.
- According to the domain theory, magnetism in materials arises at the atomic level. The current loops formed as electrons orbit an atom mean that the atom is a tiny magnetic dipole. When these ‘atomic dipoles’ are aligned together in domains, magnets are produced.
- A material is magnetically hard if it is difficult to change its domain structure.
- The domain structure of magnetically soft materials easily changes in a magnetic field.
- Permanent magnets are made from magnetically hard elements such as cobalt and alloys such as steel, alnico and magnadur. Temporary magnets are made from magnetically soft materials such as iron (e.g. wrought iron) or mumetal.
- A soft-iron core in an electromagnet increases the magnetic field strength considerably when the current is flowing. When the current is switched off, an electromagnet loses its magnetism. The magnet produced by this method is therefore called a temporary magnet.

Answers to ITQs

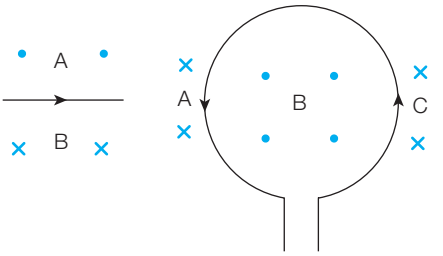
ITQ1 The magnetic force increases as the distance between the magnets decreases.

ITQ2 Arrangements 2, 3, 5, 6.

ITQ3 B

ITQ4 Towards B (applying the right-hand grip rule).

ITQ5



ITQ6 Girl, N; boy, S.

ITQ7 No. Because of the high temperatures, there will be vigorous random vibrations, and hence it is more likely that atomic dipoles will not line up to form domains.

ITQ8 The atoms of the material are probably magnetic dipoles, but these dipoles do not align together to form large domains.

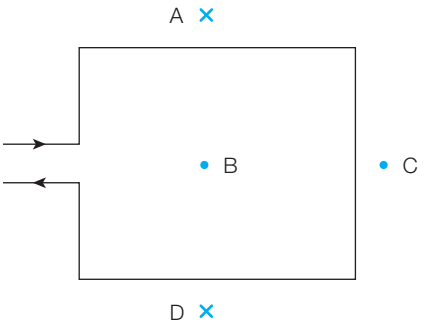
ITQ9 Either the bar is a temporary magnet, or it is a permanent magnet with its S-pole at end A.

Examination-style questions

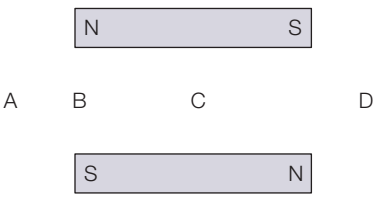
1 Under which conditions will a piece of soft iron be magnetised *most* strongly when placed in a magnetic field?

	magnetic field	temperature
A	strong	high
B	strong	low
C	weak	high
D	weak	low

2 In the diagram, which position correctly indicates the direction of the magnetic field due to the current shown in the rectangular loop of wire.



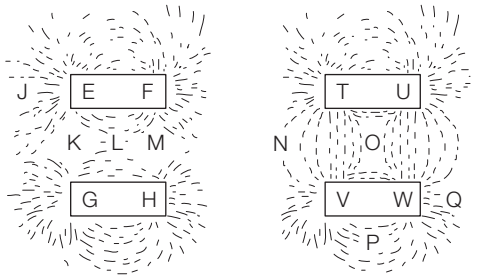
3 The diagram shows two identical permanent magnets with opposite poles facing each other.



(i) In which region is there *most* likely to be a neutral point?

(ii) In which region is there *most* likely to be a fairly uniform magnetic field?

4 In diagrams (a) and (b), magnetic field patterns between two permanent magnets have been shown using iron filings.



(i) In diagram (a), the poles could be as follows:

	north	south
A	E	G
B	G	F
C	F	H
D	H	F

(ii) In diagram (a), a neutral point is found at:

A J only
B J and K
C K and M
D L only

(iii) In diagram (b), a neutral point is found at:

A N
B O
C P
D Q

5 You are given a plotting compass and a bar magnet, both with unmarked ends.

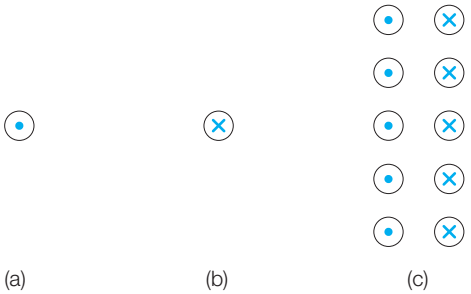
(i) Briefly describe how you would determine:

(a) which end of the compass is an N-pole;

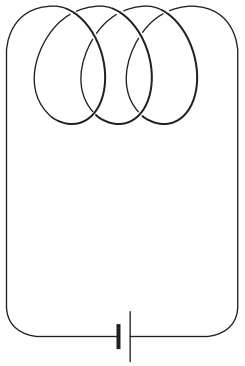
(b) which end of the bar magnet is an S-pole.

(ii) The S-pole of the bar magnet is brought near to one end B, of a soft-iron metal bar B. Draw a labelled diagram to show what magnetic field you would expect to find around the bar.

6 Sketch the field lines for the magnetic fields associated with the currents shown in the diagrams. (The dots and crosses show the directions of the currents.)



7 The diagram shows a solenoid connected to a battery. (The solenoid is an electromagnet.)



(i) Indicate the direction of the current in the wire by an arrow.

(ii) Label the S-pole of the electromagnet.

(iii) State *three* ways of making the electromagnet more powerful.

(iv) Briefly describe *one* use of electromagnets.

(v) Briefly explain why a solenoid carrying a.c. (alternating current) can be used to demagnetise a magnet.

28 Electromagnetism

By the end of this chapter you should be able to:

- describe an experiment that demonstrates the existence of a force on a current-carrying conductor placed in a magnetic field
- sketch the resulting magnetic flux pattern when a current-carrying wire is placed perpendicularly to a uniform magnetic field
- apply Fleming's left-hand (motor) rule to predict what will happen when current direction is perpendicular to a uniform magnetic field
- recall that the force on a current-carrying conductor placed perpendicular to a magnetic field depends on the strength of the field and on the magnitude of the current
- draw diagrams to show and explain the action of the following applications of electromagnetic force:
  - a simple d.c. motor
  - a moving-coil galvanometer
- describe simple experiments that demonstrate an induced e.m.f. caused by changing magnetic flux
- predict the direction of induced current, given the direction of motion of a conductor and the direction of a magnetic field
- describe simple experiments to show how the magnitude of the e.m.f. induced in a conductor depends on the rate of change of magnetic flux experienced by the conductor
- explain the action of the simple alternating current (a.c.) generator
- sketch graphs to represent the output from a simple a.c. generator
- draw a diagram of a transformer
- give a simple explanation of the principles of operation of a transformer
- describe the features of a transformer that make for efficiency
- recall that, for an ideal transformer, $P_{out} = P_{in}$
- describe experiments to show that, for an ideal transformer, $\frac{V_s}{V_p} = \frac{N_s}{N_p} = \frac{I_p}{I_s}$
- recall and apply the transformer formulas to solve simple numerical problems
- recall the advantages of using alternating current for transmission of electrical energy

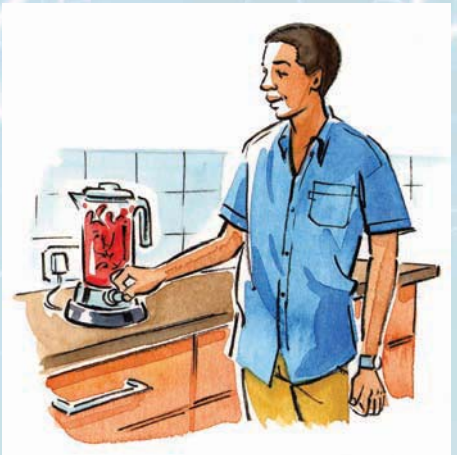


Figure 28.1 The motor in this blender makes use of electromagnetic force.

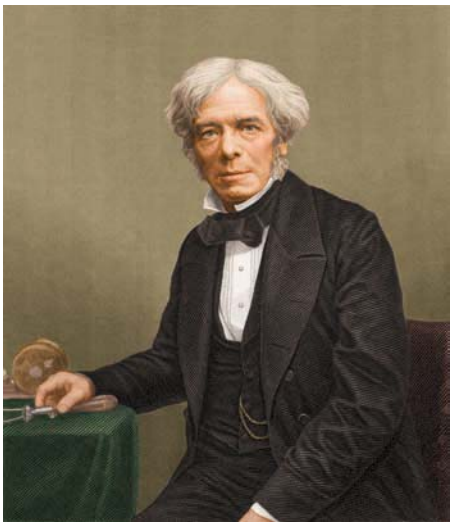
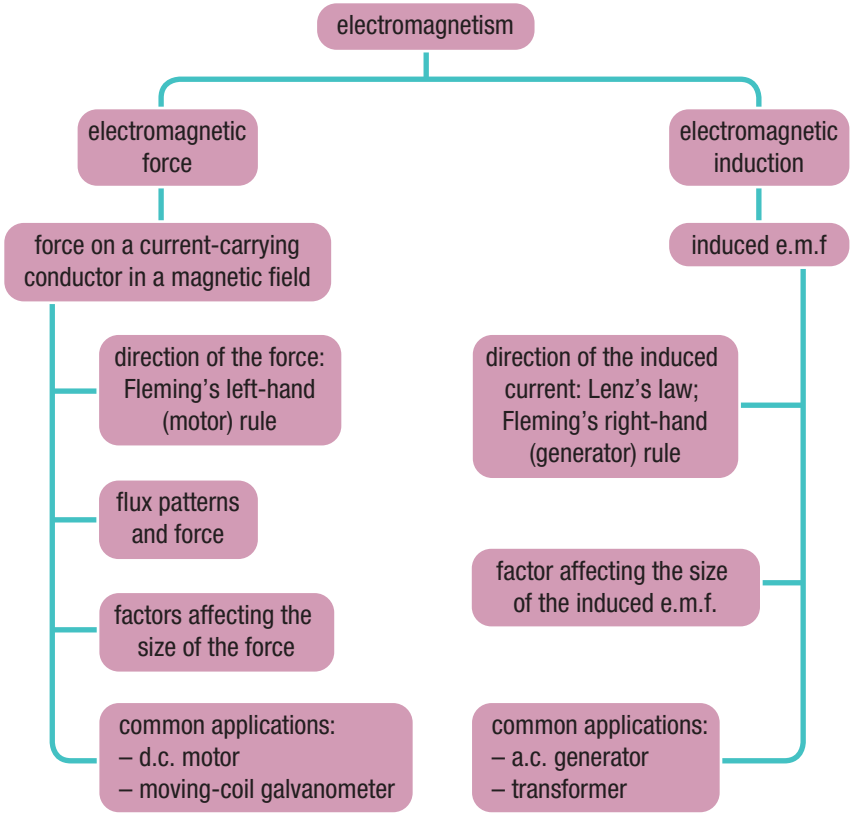


Figure 28.2 Michael Faraday (1891–1867).



Another link between electricity and magnetism

CHAPTER 27

As we saw in chapter 27, Oersted's discovery demonstrates a link between magnetism and electricity. Magnetism seems to result from current electricity, that is, from charges in motion. A magnetic field exists around a current-carrying conductor.

Another connection between magnetism and electricity was made by Michael Faraday (figure 28.2). He found that a changing magnetic field could give rise to an electric current.

Internet search term: Faraday's discovery of electromagnetic induction

Because of the discoveries of Oersted and Faraday, magnetism and electricity are treated as one phenomenon – **electromagnetism**. In this chapter, we shall focus first on electromagnetic forces, that is the forces on current-carrying conductors placed in magnetic fields. Then we shall explore **electromagnetic induction**, which is the production of current electricity from magnetism.

electromagnetism ›

electromagnetic induction ›

Electromagnetic force

Force on a current-carrying conductor in a magnetic field

We have seen that a magnetic field exists around a current-carrying conductor. Therefore, if such a conductor is brought near to a magnet or placed in another magnetic field, it will experience a force. The two magnetic fields interact, and this interaction shows itself in the **electromagnetic force**.

electromagnetic force ›

In the apparatus shown in figure 28.3, a smooth, thick wire, W, is free to roll along two smooth, conducting rails. It is placed between the N-pole and S-pole of a curved magnet. When the current is switched on in the circuit shown, wire W itself now has a magnetic field. The magnetic field of wire W interacts with that of the magnet, and wire W moves to the left. The wire moves because it experiences an electromagnetic force towards the left.

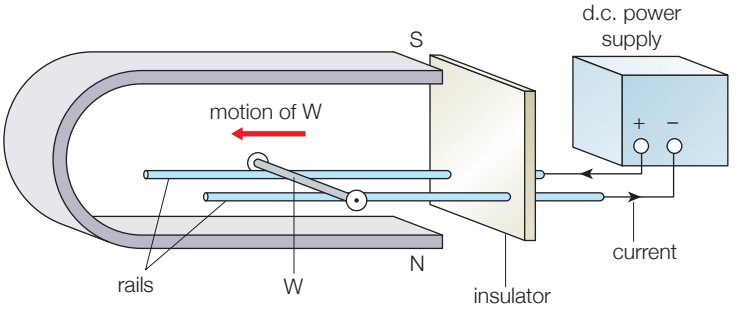


Figure 28.3 Motion of a current-carrying wire, W, in a magnetic field.

Fleming's left-hand (motor) rule

Fleming's left-hand (motor) rule ›

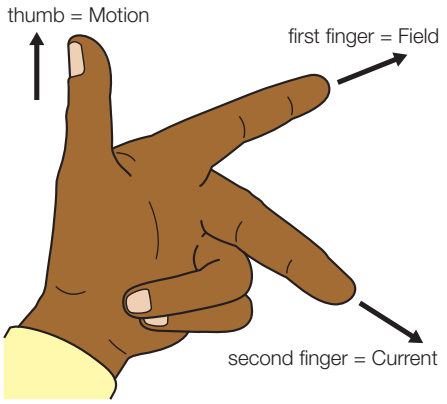


Figure 28.4 Applying Fleming's left-hand (motor) rule to figure 28.3.

The direction of an electromagnetic force can be predicted by applying **Fleming's left-hand** (or **motor**) **rule**. The thumb and first two fingers of the left hand are spread out at right angles to one another (see figure 28.4). We find that:

> **If the First finger is pointed in the direction of the Field (of the magnet), and the seCond finger is pointed in the direction of the conventional Current, the thuMb will then show the direction of the Motion produced.**

In figure 28.3, the direction of the field is upwards, and that of the current in W is out of the page. Applying Fleming's left-hand motor rule motor rule (figure 28.4), we see that the expected motion of the wire is towards the left.

Experiment shows that, if the current-carrying wire is placed parallel to a magnetic field, however, as in figure 28.5, the wire does not move. No electromagnetic force is experienced when the current is parallel to a magnetic field.

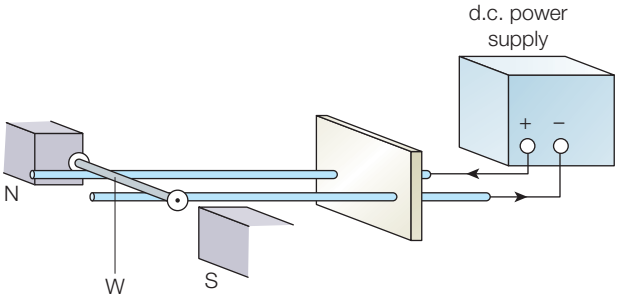


Figure 28.5 Freely moving current-carrying conductor placed parallel to a magnetic field. No force is experienced and the conductor does not move.

ITQ1
Using Fleming's left-hand (motor) rule, determine the direction of motion of the current-carrying wires in figure 28.6.

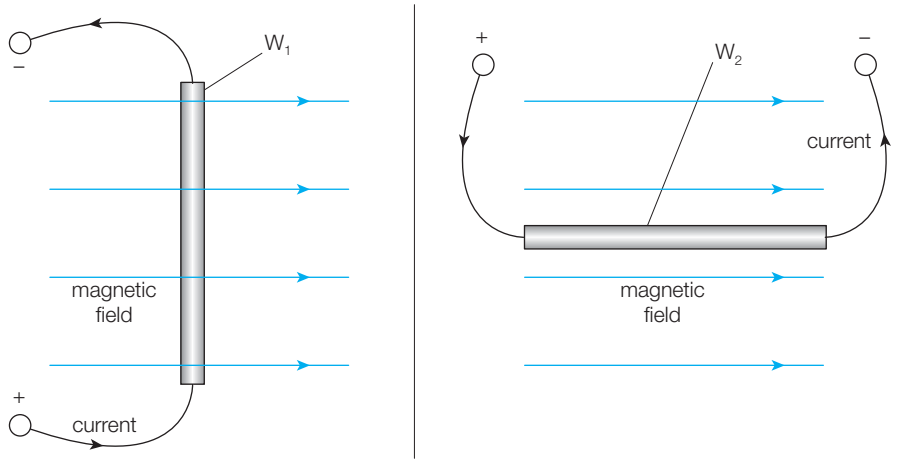


Figure 28.6

Flux patterns and electromagnetic force

magnetic flux ›

magnetic field strength ›

flux density ›

Magnetic flux is a quantity related to the strength of a magnetic field. The stronger the field, the larger the amount of lines of magnetic flux perpendicular to a given area. Thus, where magnetic lines of force are close together, the magnetic flux is greater, as shown in figure 28.7. The strength of a **magnetic field** in a region can therefore be represented as the magnetic flux per unit area (called the **flux density**) in that region. The force that a current-carrying conductor experiences when placed in the field of a magnet, or of another current-carrying conductor, is related to the flux density of the field.

Figure 28.8 (a) shows the magnetic flux pattern due to a straight, isolated, current-carrying conductor. The flux density is greatest near to the current-carrying wire.

We can use the right-hand grip rule (page 421) to determine the magnetic flux pattern due to a straight current-carrying conductor.

In figure 28.8 (b), the net magnetic flux pattern is shown for the current-carrying conductor placed at right angles to a uniform magnetic field. In region A, the flux lines are closely spaced and we say that the flux density there is higher. In region B the lines are more widely spaced and the flux density there is lower. Hence, magnetic forces would be greater at A than at B. The wire moves from A towards B (red arrow), that is *from a region of higher flux density to one of lower flux density*. At the neutral point X, the fields in opposite directions cancel.

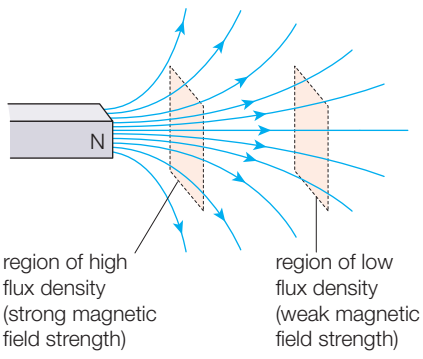
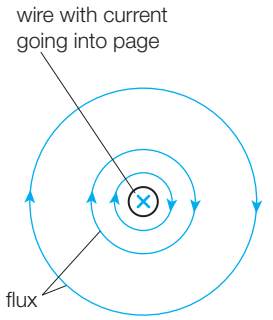
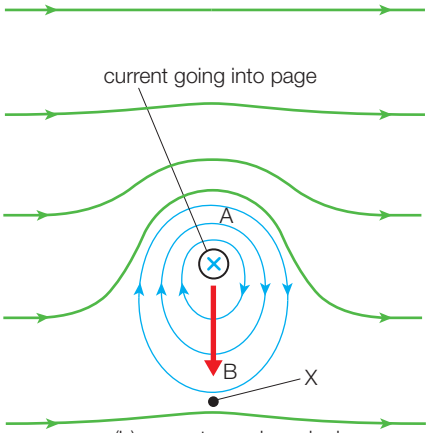


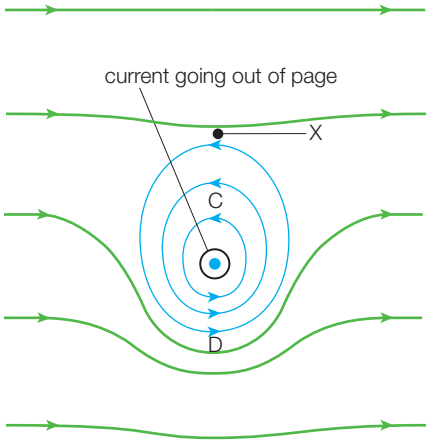
Figure 28.7 Magnetic flux density and magnetic field strength.



(a) magnetic field associated with isolated current-carrying wire



(b) current-carrying wire in uniform magnetic field



(c) current-carrying wire in uniform magnetic field

Figure 28.8 (a)–(c) Magnetic flux patterns associated with current-carrying conductors.

ITQ2

In which direction will the current-carrying wire move when placed at right angles to the uniform magnetic field as shown in figure 28.8 (c)?

Applying Fleming's left-hand (motor) rule to figure 28.8 (b), with the current going into the page and the magnetic field as shown, gives the same result for the direction of motion of the wire, that is, from A towards B.

Factors affecting the size of the electromagnetic force

The apparatus in figure 28.3 can be used to investigate factors that affect the size of the electromagnetic force. If a large current is used at switch-on, the wire W moves faster than if a smaller current is used. If the magnet shown is replaced by a more powerful magnet, wire W also moves more quickly when the same current is switched on. A wire twice the length of W also moves more quickly for the same current and magnetic field strength.

Thus, the size of the electromagnetic force depends on the following three factors:

1 the size of the current;
2 the size of the magnetic field; and
3 the length of the conductor that is in the field.

As each of these factors increases, so does the electromagnetic force.

Applications of the electromagnetic force

Simple direct current (d.c.) motor

split-ring ›
carbon brushes ›

Figure 28.9 shows the principles of operation of a simple d.c. motor. A rectangular coil is connected to two conducting half-rings, R_1 and R_2 (forming a '**split-ring**'). The current from the battery enters the coil by way of two conducting **carbon brushes**, B_1 and B_2, which rub against the split-ring as this rotates with the coil. The coil lies in a magnetic field with its axis perpendicular to the magnetic field.

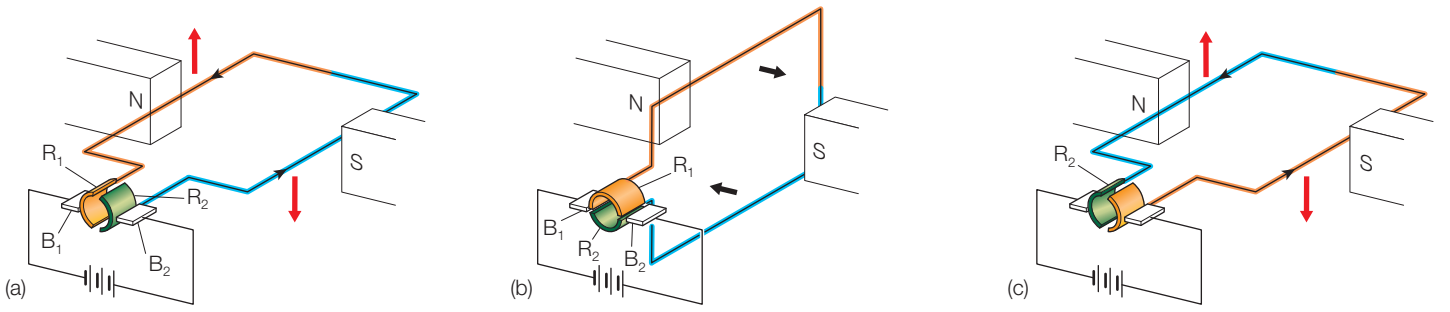


Figure 28.9 (a)–(c) How a direct current (d.c.) motor works.

When the current is switched on, the sides of the coil move as shown by the red arrows in figure 28.9 (a). (You can verify the direction of motion by applying Fleming's left-hand (motor) rule.) The coil rotates because a torque or moment (a turning effect) is produced by the electromagnetic forces in opposite directions and are at right angles to the face of the coil.

If you examine figure 28.9 (a) carefully you will notice that the current is flowing from the half of the coil marked in blue to the half marked in orange.

When the coil is at the position shown in figure 28.9 (b), no current flows momentarily because the brushes are over the gaps of the split-ring. However, the momentum of the coil keeps it turning in the directions of the black arrows shown. If the circuit was switched on when the coil was at the position shown

Carbon brushes: The carbon brushes in the motor are usually made of graphite, a slippery form of carbon. The graphite is softer than the metal (usually copper) of the split-rings, and wears away faster than the copper. The worn brushes are easily replaced at low cost. It is important that the brushes are able to withstand the very high temperatures that develop in motors.

in (b), there would be no rotation. The gaps would prevent current from entering the coil, but even if there was a current, the coil would still not rotate as it would experience no torque. This is because the forces, though opposite in direction, would be parallel to the plane of the coil, that is, directed upward and downward (not shown in the diagram).

As the coil rotates, the brushes again establish contact with the split-ring, as shown in figure 28.9 (c). However, as you can see, the current will now flow from the orange half of the coil to the blue half: the direction of the current through the coil itself has been reversed. Applying Fleming's left-hand rule, we see that forces producing rotation in the same direction as in (a) are exerted on the sides of the coil and it continues to move in the direction.

Why must the direction of current in the coil be reversed? Suppose the current direction in figure 28.9 (c) was still from the blue half of the coil to the orange half, as in figure 28.9 (a). Applying Fleming's left-hand rule, you can see that the coil would start to rotate in the opposite direction. So to keep the coil rotating in one direction we need to change the direction of the current once every half-turn. The split-ring arrangement ensures that the current is reversed in the coil at the right moment. Since the split-ring reverses the current, it is also called a split-ring **commutator**.

commutator ›

Figures 28.10 and 28.11 show the parts of a practical d.c. motor:

- A permanent field magnet (figure 28.10). Sometimes a field electromagnet, called a stator (figure 28.11), is used instead of the permanent magnet shown in figure 28.10.
- A split-ring (figure 28.10) or segmented (figure 28.11) commutator, which reverses the current in the coil.
- An **armature**, which consists of the coil wrapped around a soft-iron core. The soft-iron core increases the field strength due to the motor coil, so that a large turning moment (torque) is produced.

armature ›

ITQ3

What is the name of the part labelled X in figure 28.10?

SBA activity 28.1 involves using a d.c. motor.

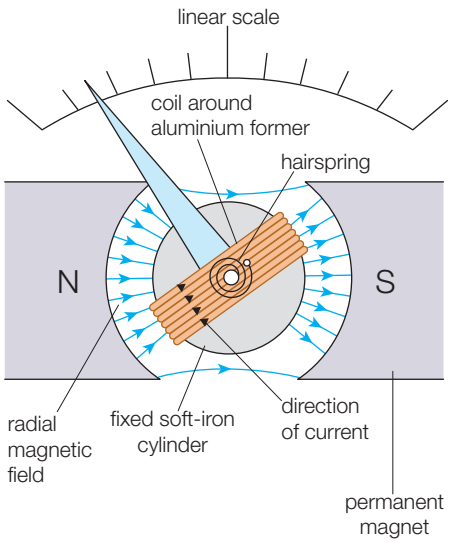


Figure 28.12 A moving-coil centre-zero galvanometer. Note that the direction of the current determines the direction in which the coil would rotate. The current goes into the page on the right side of the coil and out of the page on the left side.

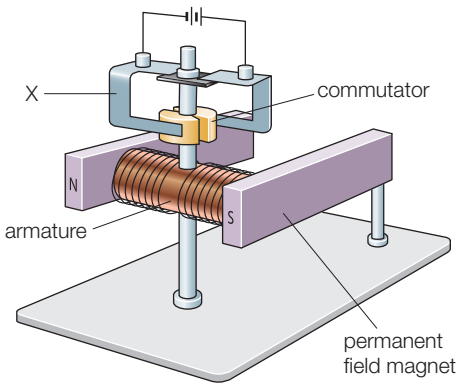


Figure 28.10 A practical d.c. motor.

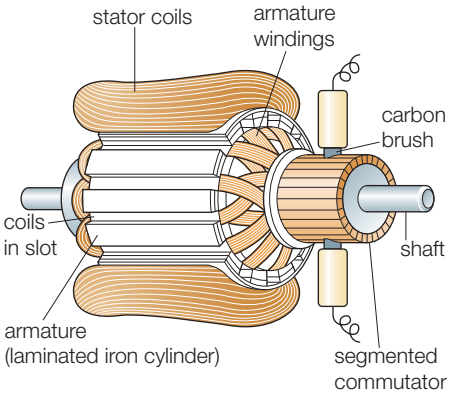


Figure 28.11 A motor armature with several coils and commutators. This arrangement helps the motor to run smoothly, since equal torques are given to the armature several times during one cycle.

Moving-coil galvanometer

A curved permanent magnet and a fixed, cylindrical soft-iron core (see figure 28.12) provide a radial magnetic field in this instrument. A rectangular coil, wound on an aluminium frame former that is fixed over the soft-iron core, is pivoted so that the coil can move freely in the magnetic field. Current enters and leaves the coil by two controlling springs.

When there is a current in the coil, an electromagnetic force produces an electromagnetic couple (a moment produced by two equal but opposite forces on the sides of the coil), which causes the coil (and attached needle)

to rotate. The controlling springs now come into play and bring the coil to a stop at a position where the electromagnetic couple (which is proportional to the current in the coil) just balances the restoring couple provided by the springs. It can be shown from the fact that the two opposing couples balance at equilibrium that the angle of deflection of the coil is proportional to the current in the coil. This is an important relationship, as it tells us why the scale of a moving-coil galvanometer is linear.

MATHEMATICS: direct proportion between two variables

The radial field mentioned earlier ensures that the plane of the coil is always parallel to the magnetic field in which the coil finds itself. This leads to the linear relationship between the deflection of the coil and the current, and hence the uniform linear scale mentioned above. The springs return the coil and attached needle to zero when the current becomes zero in the coil and exert a restoring couple on the coil while there is a current.

The moving-coil galvanometer can be made very sensitive (that is, able to detect tiny current) by using:

1 a large number of turns in the coil;
2 a weak spring;
3 a powerful magnet; and
4 a coil of large surface area.

IT:Q4

Apply Fleming's left-hand (motor) rule to determine the direction of rotation of the coil in figure 28.12 for the direction of current shown.

Electromagnetic induction

Faraday's experiments

As mentioned earlier, Michael Faraday (and indeed others) showed that electricity could be produced from magnetism.

In one of Faraday's experiments, a coil of copper wire wrapped around a paper tube was connected to a sensitive galvanometer (figure 28.13). When a magnet was thrust into the coil, the galvanometer needle showed a momentary deflection. When the magnet was pulled out of the boil, the needle again showed a momentary deflection, this time in the opposite direction. No deflection was observed if the magnet was kept stationary outside or inside the coil.

Remember that something 'induced' is always brought about by the influence of something else. Do you remember how we can 'induce' a charge by the influence of another charge?

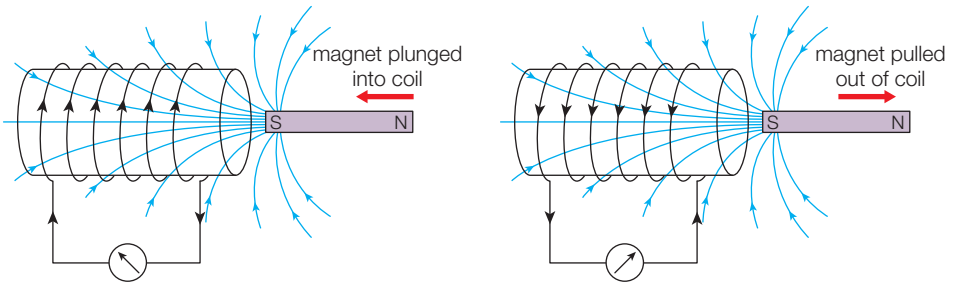


Figure 28.13 Electromagnetic induction.

The deflection of the galvanometer needle showed that electric current had been induced in the copper wire. An electromotive force (e.m.f.) must have been **induced** in the copper wire circuit resulting in a current. Faraday had demonstrated that current electricity could be produced from magnetism. From further experimentation with coils and magnets, Faraday found that the size of the **induced current** depended on:

induced current ›

induced e.m.f. ›

1 the number of turns in the coil;
2 the area of the coil;
3 the strength of the magnet; and
4 the speed with which the magnet was thrust into (or pulled out of) the coil.

Magnetic flux and the induced e.m.f.

ITQ5

One end of a magnet is held stationary and perpendicular to the face of a coil as shown in figure 28.15. If the coil is quickly rotated through 90° about a vertical axis, AB, will there be an induced e.m.f. in the coil? Explain your answer.

magnetic flux linkage ›

Faraday's law ›

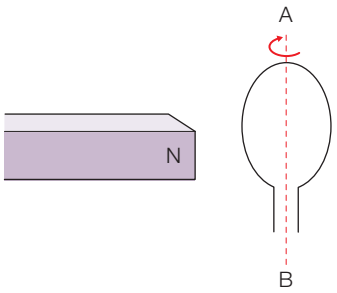


Figure 28.15

Magnetic flux is a measure of the number of magnetic field lines passing through an area perpendicular to the field lines. Figure 28.14 shows that there is no flux through a coil whose plane is *parallel* to the magnetic field. There is maximum flux through the face of a coil whose plane is *perpendicular* to the magnetic field.

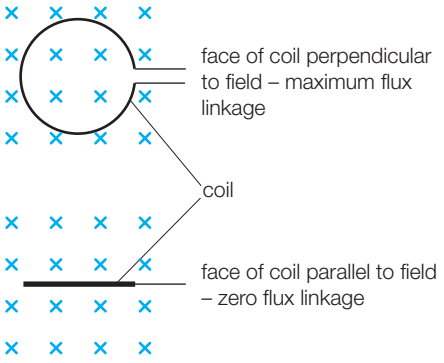


Figure 28.14 Magnetic flux linking a coil.

The amount of flux passing through the face of a coil is called the **flux** 'linking' the coil (i.e. the **magnetic flux linkage**). Faraday concluded that an e.m.f. is induced whenever there is a change in the flux linking a coil. The greater the change in the amount of flux and the more quickly it happened, the greater the e.m.f. induced. **Faraday's law** of electromagnetic induction states the following:

> **The size of the e.m.f. induced in a circuit is directly proportional to the rate of change of magnetic flux linking the circuit.**

Factors affecting the size of the induced e.m.f.

As indicated in the preceding paragraphs, there are several factors that affect the size of the induced current, and hence the induced e.m.f., in a coil. In the following activity, you will investigate these factors.

Practical activity 28.1

Factors affecting the size of the induced e.m.f.

An apparatus similar to figure 28.13 can be used in these activities. However, a centre-reading millivoltmeter can be used instead of the centre-reading galvanometer shown to measure the induced voltage directly.

You are to plan and design activities to investigate how the following affect the magnitude of the e.m.f. induced in a coil (note: you will need to pay attention to control of variables):

1. increasing the number of turns of the coil;
2. increasing the surface area of the coil, to allow as much flux as possible through the coil;
3. increasing the magnetic flux (by using a more powerful magnet or electromagnet, or by placing a soft-iron core in the coil); and
4. causing a more rapid change of flux through the coil (by moving the magnet or the coil more quickly towards or away from each other).

Experiments have shown that all four factors listed above result in increasing the induced e.m.f.

The direction of the induced current: Lenz's law

When the N-pole of a magnet is moved towards a coil, an anticlockwise current appears in the coil (figure 28.16). This current can be detected by a sensitive, centre-zero galvanometer. Applying the 'AN' rule (page 422), we see that an N-pole is created at the end of the coil, facing the incoming N-pole, as

shown in figure 28.16. Since like poles repel, work must be done against this repulsive force in order to establish the induced current.

When the N-pole is withdrawn from the coil, a clockwise current appears in the coil. Using the 'CS' rule (page 422), we see that an S-pole is created at the end of the coil. Since unlike poles attract, the motion of the magnet is again opposed. A series of such experiments led Heinrich Lenz to formulate what is now known as **Lenz's law** of electromagnetic induction:

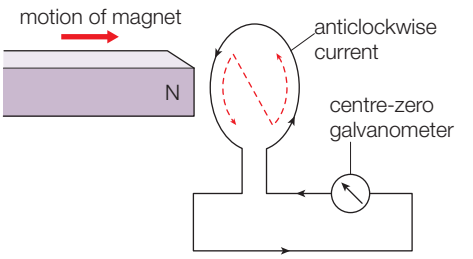


Figure 28.16 Illustrating Lenz's law.

Lenz's law ›

> **The direction of the induced current is such as to oppose the change giving rise to it.**

Practical activity 28.2

Investigating Lenz's law

An apparatus similar to figure 28.16 can be used in these activities.

Method

1 First, make sure you know the direction of the current through the meter when the centre-zero galvanometer deflects to the left (and to the right). (You may need to test the galvanometer with a small current, e.g. using a 1.5V cell and a 1000 Ω in series with the meter in order to find out this – but first check with your teacher before doing this test!)

2 Connect up the circuit of figure 28.16 and investigate the direction of the induced current in the coil when:

(a) the N-pole of the magnet is thrust into/out of the coil;

(b) the S-pole of the magnet is thrust into/out of the coil.

Question

Are your observations in agreement with Lenz's law? Explain your conclusion.

ITQ6

What is the direction of the current in the coil in figure 28.16 if an S-pole of a magnet: (i) remains steady within the coil; (ii) is moved away from the coil?

Worked example 28.1: Electrical energy obtained during electromagnetic induction

There is energy in an electric current. From what source does the current induced in a coil obtain its energy?

Answer

The work done against the force that opposes the motion of the magnet (that is, against the change of magnetic flux linking the coil) becomes the electrical energy obtained by the coil. This is an example of the law of conservation of energy. The work done represents the energy being converted from one form – that is, the energy supplied in moving the magnet against the opposing force – to another form – that is, electrical energy.

Electromagnetic induction in a metal sheet

As an N-pole is moved towards a conducting sheet (figure 28.17), an anticlockwise current flows on the surface of the sheet so as to create an opposing magnetic pole (an example of Lenz's law). As the magnet is moved away, the current flows in the opposite direction, again producing an opposing magnetic field.

eddy current ›

The currents flowing in the sheet are known as **eddy currents**, as their flow is reminiscent of the movement of eddies in water, that is, round and round in circles. These currents can produce a heating effect in the sheet because the current is flowing through material that has resistance. Such heating is called '**Joule heating**'. If the sheet is cut into strips separated by layers of insulation, as in figure 28.18, this will prevent the formation of large eddy currents. This technique for reducing eddy currents is applied in the construction of transformer soft-iron cores (page 445) in order to reduce 'Joule heating' losses within the cores.

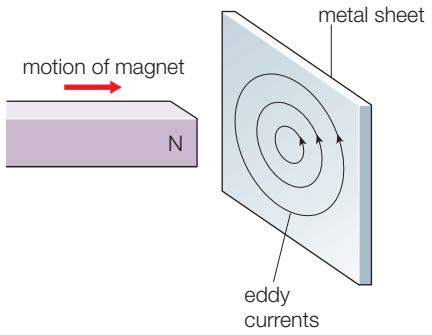


Figure 28.17 Producing eddy currents in a metal sheet.

ITQ7

What would be the direction of the current in the sheet metal in figure 28.17 if an S-pole was being moved towards the sheet?

Joule heating ›

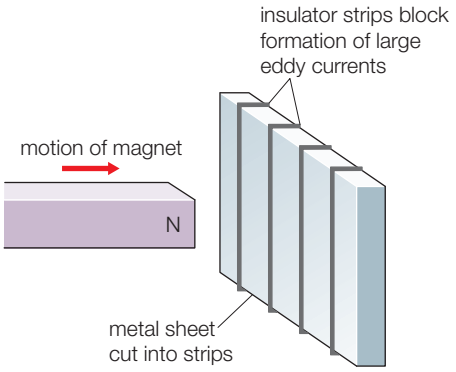


Figure 28.18 Reducing eddy current formation.

Fleming's right-hand (generator) rule ›

Electromagnetic induction in a straight wire

In one of his many experiments, Faraday found that, when a straight wire cuts across magnetic lines of flux, an e.m.f. is also induced in the wire. The direction of the induced current is given by **Fleming's right-hand (generator) rule**. The thumb and first two fingers of the right hand are spread out at right angles to one another as shown in figure 28.19. We find that:

> **If the thuMb gives the direction of Motion of the wire, and the First finger gives the direction of the magnetic Field, then the seCond finger will give the direction of the induced (conventional) Current.**

Note that Fleming's motor rule is the 'left-hand rule' but that the generator rule is the 'right-hand rule'.

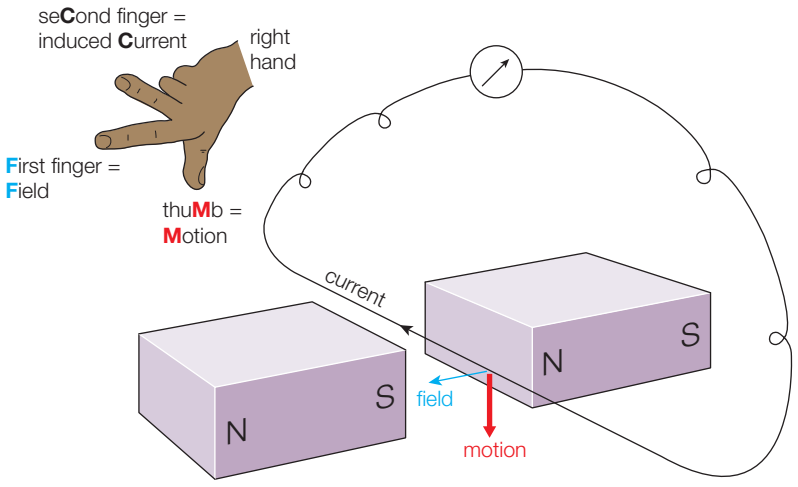


Figure 28.19 Electromagnetic induction in a straight wire and application of Fleming's right-hand (generator) rule.

Did you know that an induced e.m.f. appears across the wings of an aircraft as they cut across the Earth's magnetic flux (figure 28.20)?

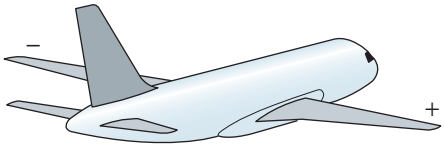


Figure 28.20

Applications of electromagnetic induction

The alternating current generator or alternator

generator ›

The principle of operation of a simple alternating current (a.c.) **generator** is shown in figure 28.21. The ends of a rotating rectangular coil, shown as abcd, are connected to metal *slip-rings*. Carbon *brushes*, rubbing against the slip-rings, conduct the induced current away from the coil. The brush and slip-rings arrangement is needed to avoid the output wires fixed to the ends of the coil becoming intertwined as the coil rotates.

We can apply Fleming's right-hand (generator) rule to determine the direction of the output current, and hence the polarity of the output e.m.f. or voltage of the rotating coil (figure 28.22).

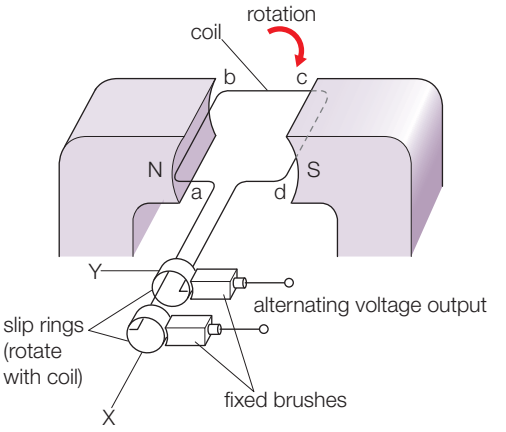


Figure 28.21 A simple alternating current (a.c.) generator. The rotating coil is shown in the horizontal position.

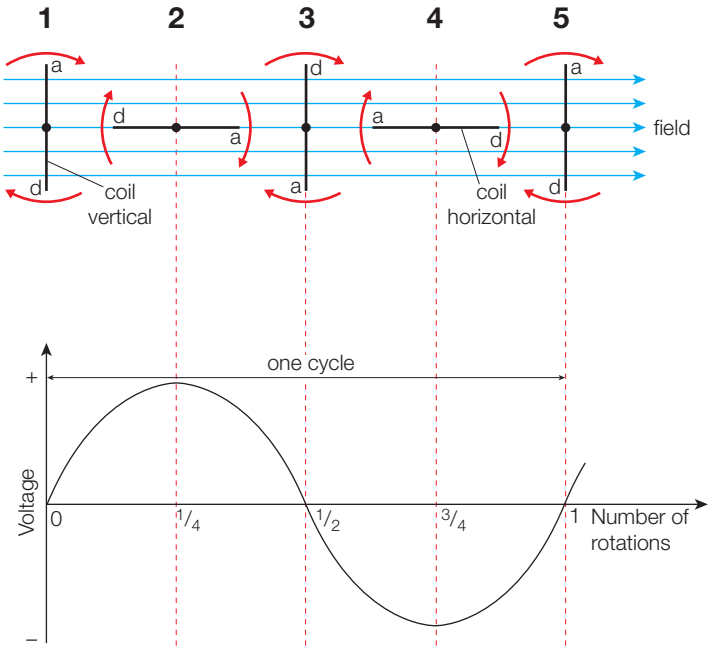


Figure 28.22 Orientation of coil and corresponding output voltage of generator coil during one cycle.

1 *Coil vertical*. Sides ab and cd of the coil are both moving parallel to the field. No flux lines are being cut by ab and cd. There is no induced e.m.f. and hence no induced current.

2 *Coil horizontal (one quarter-cycle later)*. The side ab of the coil is now moving downwards on the right, and side dc is moving upwards on the left, and they are both cutting across magnetic lines of flux at right angles. A maximum e.m.f. is induced, and current flows through the coil. Applying Fleming's right-hand rule to side ab and then to side dc, you will find that

the direction of the induced current through the coil is towards slip-ring X (figure 28.21).

3 *Coil vertical again (one quarter-cycle later)*. The induced e.m.f. is again zero.

4 *Coil horizontal again (one quarter-cycle later)*. Side ab of the coil is now moving upwards on the left, and side dc is moving downwards, and again they are both cutting across magnetic lines of flux at right angles. A maximum e.m.f. is again induced. However, applying Fleming's right-hand rule to ab and cd in turn, you will find that the direction of the induced current is now through the coil towards slip-ring Y.

MATHEMATICS: sine graph

A graph of the output voltage for one complete rotation (cycle) of the coil is shown in figure 28.22.

Power station generators usually have a rotating electromagnet (called a **rotor**) surrounded by a stationary coil called a **stator** (figure 28.23). Large voltages are obtained directly from the stator. If the slip-ring and brushes arrangement were used, the brushes would wear away extremely quickly on account of the large currents between the brushes and the slip-rings. There would also be considerable friction between the two, causing further wear on the brushes.

rotor ›
stator ›

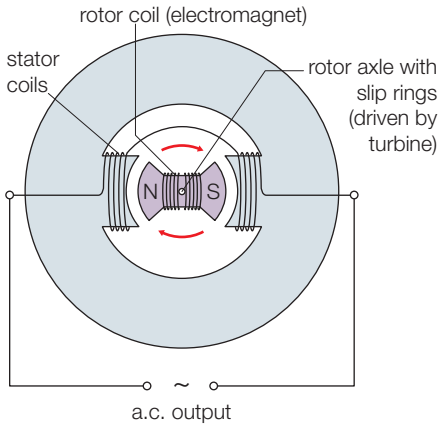


Figure 28.23 Alternator with stator output.

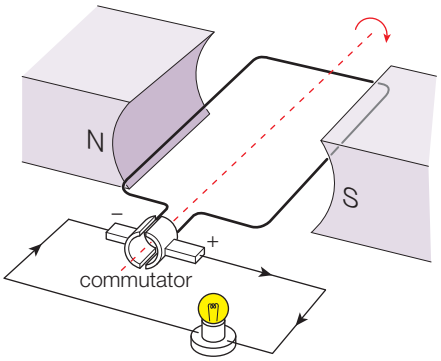


Figure 28.24 A simple d.c. generator.

The direct-current generator or dynamo

A simple direct-current (d.c.) generator or dynamo is shown in figure 28.24. A commutator is used instead of slip-rings so that the direction of the current induced in the coil is reversed every half-cycle. Application of Fleming's right-hand (generator) rule shows that the side of the commutator to the right will always have a positive polarity for the direction of rotation of the coil indicated.

Figure 28.25 (a) shows the variation in output voltage with time for one such coil. Figure 28.25 (b) shows the much smoother output voltage obtained if the dynamo rotor has many coils whose planes are at equal angles to one another, each coil being connected to a separate commutator segment.

ITQ8

Which device does the d.c. dynamo (figure 28.24) resemble in its construction? Can this device function as a d.c. dynamo?

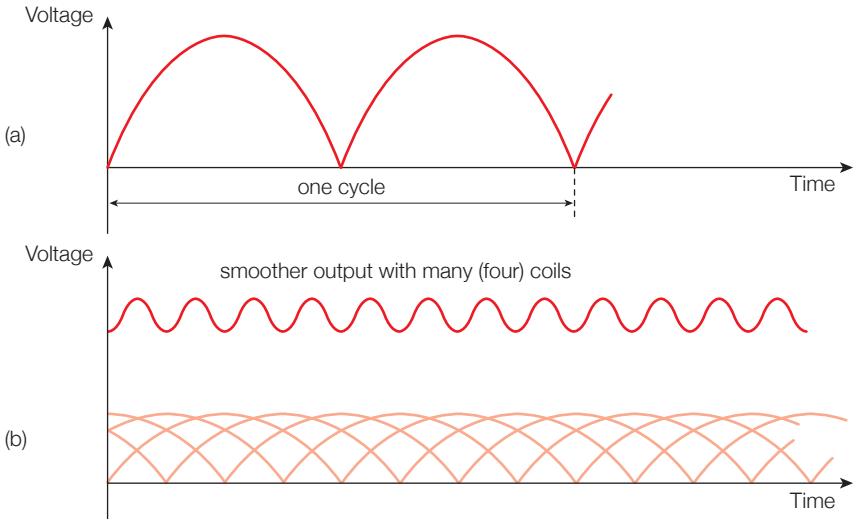


Figure 28.25 (a) Voltage vs. time graph for output of a simple d.c. dynamo (b) Smoother output voltage obtained using four coils.

The transformer

Mutual induction

primary coil | secondary coil ›

In another experiment, Faraday used a coil connected to a battery, in place of a permanent magnet, to induce an e.m.f. in a second coil. Both coils were wrapped around an iron ring as shown in figure 28.26. The coil in the battery circuit was called the **primary coil** and the other was the **secondary coil**. This coil was connected to a centre-zero galvanometer.

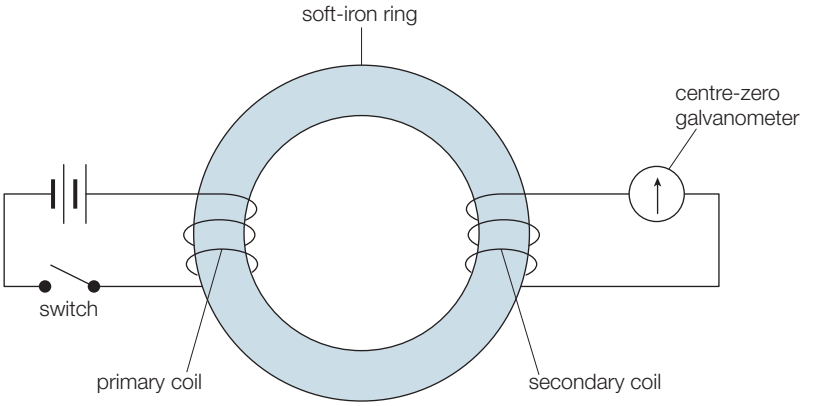


Figure 28.26 Faraday's demonstration of mutual induction.

When the switch was closed, a current appeared momentarily in the secondary coil, indicating the presence of an induced e.m.f. While the current remained steady in the primary coil, there was no current in the secondary coil. When the switch was opened, a momentary current flowed in the opposite direction. The changing flux from the primary coil had caused an induced e.m.f. in the secondary coil.

This phenomenon in which an e.m.f. is induced in a coil by flux change from another coil became known as **mutual induction**.

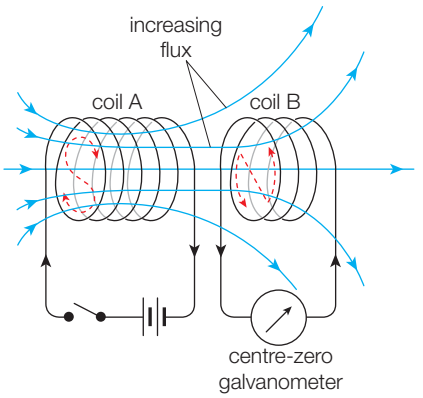


Figure 28.27

mutual induction ›

The presence of the iron in figure 28.26 increased the magnetic flux, as the magnetic field produced by the current-carrying primary coil caused the domains within the iron to become aligned with the field (see page 425). The ring shape ensured that most of the flux produced by the primary coil linked the secondary coil. When the current was switched off, the magnetic field due to the primary coil returned to zero. The domains within the iron also returned to their random state of orientation, resulting in practically zero flux as well.

Lenz's law also applies to mutual induction. When the switch in figure 28.27 is closed, the momentary current induced in coil B is anticlockwise. The N-pole formed as a result in coil B repels the changing N-pole of A, that is, it opposes the inducing flux.

Practical activity 28.3

Caution: To be done only under close supervision of your teacher.



Mutual induction

Using apparatus similar to figure 28.27, find out how the magnitude of the current induced in coil B is affected by:

1 the number of turns in coil A;
2 the number of turns in coil B.

(Remember to control variables.)

The principle of the transformer

If a continuously changing e.m.f. is applied to a primary coil, the current in the coil, and therefore the magnetic flux it produces, will also change continuously. An induced e.m.f. will be produced in a nearby secondary coil, through the process of mutual induction, as the continually changing flux from the primary coil cuts cross the secondary coil. This is the principle of operation of a **transformer**. If there are no energy losses between the primary and secondary coils, it can be shown that

transformer ›

$$\frac{V_p}{V_s} = \frac{N_p}{N_s}$$

where V_p and V_s are the voltages (e.m.f.s) across the primary (p) and secondary (s) coils, and N_p and N_s are the numbers of turns in the primary and secondary coils (figure 28.28). That equation shows that the **output voltage** appearing across the secondary coil is proportional to the number of turns in the coil.

output voltage ›

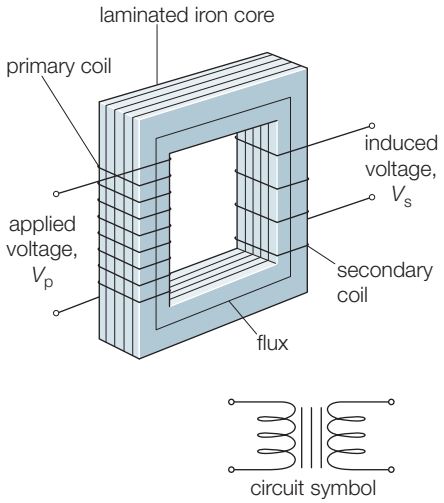


Figure 28.28 A simple transformer. Note the symbol for the transformer; the primary is always drawn on the left and the secondary on the right. The lines (1 to 4) between these two represent the laminated core.

Worked example 28.2: Obtaining a low voltage from a high voltage

Assuming no power loses, calculate the voltage V_s appearing across the secondary coil in the transformer in figure 28.28 when 40 V is applied across the primary coil.

Solution

Primary voltage, $V_p = 40\,\text{V}$

Secondary voltage, $V_s = ?$

Number of turns in primary coil, $N_p = 8$

Number of turns in secondary coil, $N_s = 3$

Using

$$\frac{V_p}{V_s} = \frac{N_p}{N_s}$$

we get

$$\frac{V_s}{N_p} = V_p \times N_s$$

$$= 40 \times \frac{3}{8}$$

$$= 15\,\text{V}$$

input voltage ›

step-down transformer ›

step-up transformer ›

A transformer is therefore able to change a given alternating **input voltage** to a higher or a lower voltage. The transformer shown in figure 28.28 is called a **step-down transformer**, since the output voltage is less than the input voltage; the number of turns in the secondary coil is less than the number of turns in the primary coil. The output voltage in a **step-up transformer** is greater than the input voltage; here the number of turns in the secondary coil is greater than the number of turns in the primary coil.

The output current from a transformer can be estimated using the formula for power, $P = IV$, where P is the power within a coil for a current I in the coil and a voltage V applied across the coil.

Worked example 28.3: Transformer current and power

A transformer is used with the 110 V a.c. supply to deliver 9.0 A at 8.0 V to a heating coil.

(i) Draw a circuit diagram to illustrate the above information.

(ii) Calculate the power delivered by the secondary coil.

(iii) Calculate the current in the primary coil.

Solution

(i)

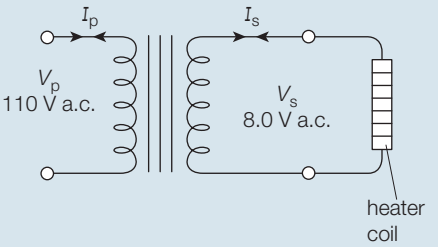


Figure 28.29 A transformer providing a low voltage (and high current) to a heating coil.

(ii) In the secondary coil:

current, $I_s = 9.0\,A$

voltage, $V_s = 8.0\,V$

So if the power in the secondary coil is P_s, then

$P_s = I_s V_s$

$= 9.0\,A \times 8.0\,V$

$= 72\,W$

(iii) Assuming no power losses, the power delivered by the secondary coil, P_s, is equal to the power, P_p, received from the primary coil, i.e. $P_s = P_p$.

In the primary coil:

voltage, $V_p = 110\,V$

current, $I_p = ?$

The powers are

$P_s = P_p = I_p V_p$

$72\,W = I_p \times 110\,V$

$I_p = 0.65\,A$

The current in the primary circuit is 0.65 A.

Worked example 28.3 shows that, as the output voltage of a transformer is stepped down, the output current is stepped up.

IT Q9

What would happen to the output current from a transformer if the output voltage was stepped up?

The equation relating current to voltage in transformer coils can be obtained from power considerations. For an ideal transformer, the power output, P_s, at the secondary coil is equal to the power input, P_p, at the primary coil. Thus

$$I_p V_p = I_s V_s$$

or

$$\frac{V_p}{V_s} = \frac{I_s}{I_p}$$

where I_p and I_s are the currents in the primary and secondary coils, respectively. The first equation suggests that, with the aid of a transformer, a fixed quantity of electrical power could be sent at low voltage and high current or at high voltage and low current.

Worked example 28.4: Sending power from a generating station

The output voltage from a particular generating station can be stepped up to 10 000 V (10 kV) or to 100 000 V (100 kV) using a transformer. Would it be better to send 1 MW of power from the generating station at 10 kV or at 100 kV? Assume that the power distribution lines have a total effective resistance of 2 Ω.

The formulas for power, $P = IV = I^2R$ were discussed in chapter 26 (page 412)

Solution

We make use of the equation for power: $P = IV$.

At 10 000 V, we get

$$\frac{I_1}{V_1} = P_1$$
$$= \frac{1\,000\,000\text{ W}}{10\,000\text{ V}}$$
$$= 100\text{ A}$$

In lines of resistance, R, the power loss through heating is given by $P = I^2R$.

Thus, power loss, $P_1 = I_1^2R = 100^2 \times 2 = 20\,000$ W.

At 100 000 V, we get

$$\frac{I_2}{V_2} = P_2$$
$$= \frac{1\,000\,000\text{ W}}{100\,000\text{ V}}$$
$$= 10\text{ A}$$

Thus, power loss, $P_2 = I_2^2R = 10^2 \times 2 = 200$ W.

While power transmission at higher voltages reduces power loss, the wires require very good insulation to prevent current 'leaking' to the ground.

We see therefore that *much less* power is wasted if it is transmitted at very high rather than lower voltages.

Practical transformers

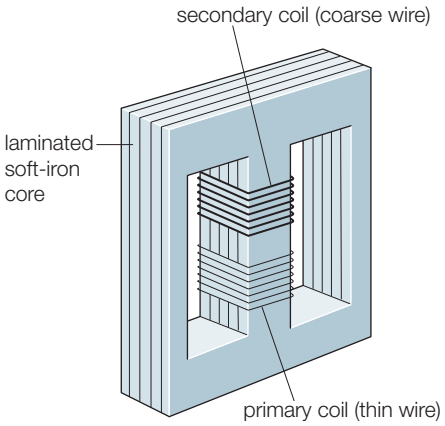


Figure 28.30 A practical transformer.

Practical transformers are carefully designed to minimise loss of power between the primary and secondary coils. This is done in the following ways:

1 Wires of low resistance (coarse wire) are used to reduce ohmic (I^2R) losses in the coil windings. Generally, coarser wires are used for the high-current (low-voltage) windings. It should be noted, however, that the use of thicker wire would increase the cost and make the transformer heavier, which could be a disadvantage.

2 *A* soft-iron core shaped into a closed path is used, to reduce magnetic flux losses between the two coils. Magnetic flux travels easily along the soft-iron path and hence nearly all the flux that leaves the primary coil passes through (or links) the secondary coil (figure 28.30)

3 A laminated core with insulated layers is used, to reduce the size of the eddy currents that will flow in the core. This will therefore reduce eddy current energy losses from Joule heating in the core (see pages 438–439).

4 Good-quality soft iron is used, which will reduce the amount of heat generated in the iron core. Every time the primary current is taken through one cycle around the core, a certain amount of heat is generated in the

core, depending on the type of soft iron used. This type of energy loss is known as hysteresis loss.

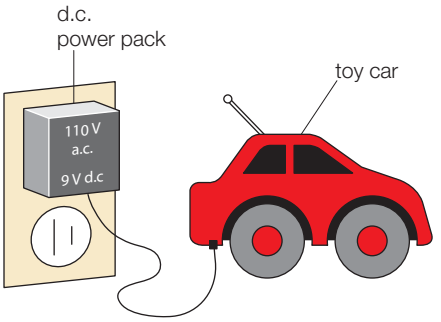


Figure 28.31 Many low-voltage d.c. power packs contain transformers to step down the voltage and rectifiers (diodes) to convert the alternating current to direct current.

Advantages of using alternating current for energy transfer

A big advantage of using a.c. for energy transfer is that it is very easy to step up or step down alternating voltages (and currents) using transformers. Power can be sent at high voltage and low current to reduce power loss (see Worked example 28.4), or at low voltage and high current as needed. Using a suitable transformer, it is easy to obtain the required voltages (or currents) from an alternating supply.

Most electronic equipment today (e.g. calculators, computers) require various voltages but use direct (d.c.) current for their operation. It is not only very easy to obtain these voltages from an a.c. voltage supply, but also very easy to convert alternating current to direct current if this is required, by using appropriate rectifiers (diodes) connected to transformers (figure 28.31).

Chapter summary

- A current-carrying conductor experiences an electromagnetic force when placed in a magnetic field.
- The magnitude of the force is proportional to the size of the current and the strength of the magnetic field.
- If the current flows at right angles to the magnetic field, the direction of the force can be predicted by applying Fleming's left-hand or motor rule.
- Electric motors and moving-coil galvanometers make use of the electromagnetic force.
- Magnetic flux is a basic property associated with magnets and current-carrying conductors.
- A measure of the size of magnetic flux is the force a current-carrying conductor experiences when placed in an area where there is magnetic flux. We can judge the relative amount of flux in an area of a magnetic field from the closeness of the lines of force in that region.
- An e.m.f. is induced in a conductor by a changing magnetic flux. The greater the rate of change of the flux, the greater the induced e.m.f.
- Lenz's law states that induced current flows in such a way as to create a magnetic field that tends to oppose the change brought about by the inducing flux.
- The direction of the current induced in a straight wire can be predicted by applying Fleming's right-hand or generator rule.
- Generators and transformers make use of electromagnetic induction for their operation.
- For an ideal transformer, $P_{out} = P_{in}$.
- For an ideal transformer,
$$\frac{V_s}{V_p} = \frac{N_s}{N_p} = \frac{I_p}{I_s}$$

Answers to ITQs

ITQ1 W_1 would move as though into the page. W_2 would not move since the current-carrying wire is parallel to the field.

ITQ2 Towards C because the field at D is stronger than that at C, as indicated by the larger number of flux lines.

ITQ3 Brush.
ITQ4 The coil will rotate clockwise.
ITQ5 Yes, because of the change in flux linkage from maximum (field perpendicular to face of the coil) to zero (field parallel to face of the coil).
ITQ6 **(i)** No current. **(ii)** Anticlockwise (an N-pole would be formed to oppose the S-pole from being moved away from the coil).
ITQ7 Clockwise (to form an S-pole to oppose the S-pole of the magnet that is moving towards the sheet).
ITQ8 The d.c. motor. Yes, since they are practically identical in construction.
!TQ9 The output current would be stepped down, since output power, given by secondary (output) current × secondary (output) voltage (or $I_s \times V_s$), remains constant.

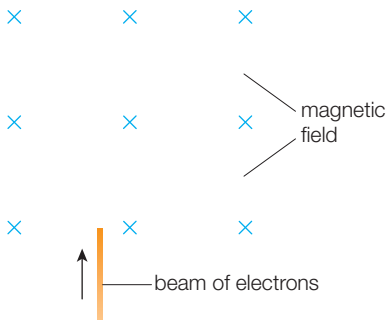
Examination-style questions

1 Which of the following parts may be found in a direct current (d.c.) motor?

I brushes
II slip-rings
III soft-iron armature

A I only
B I and II only
C I and III only
D II only

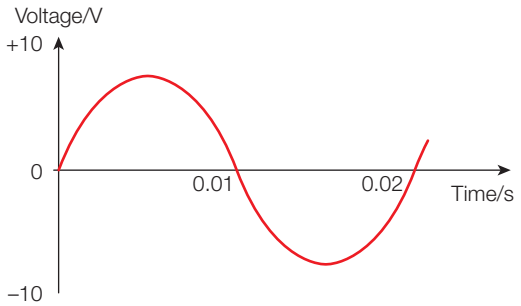
2 A beam of electrons enters a magnetic field as shown in the diagram.



In which of the following directions is the beam deflected?

A →
B ←
C ↑
D ↓

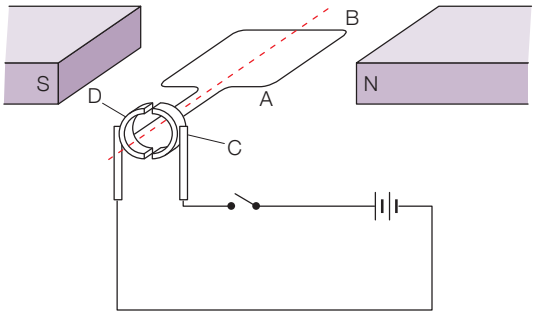
3 The graph shows the variation of an alternating voltage with time.



The frequency of the voltage is:

A 0.01 Hz
B 0.02 Hz
C 50 Hz
D 100 Hz

4 The diagram show a coil placed in the magnetic field of a certain device.



(i) In which direction will AB move when the current is switched on?

(ii) What is the name given to C?

(iii) What is the function of D?

(iv) What is the name given to the device shown?

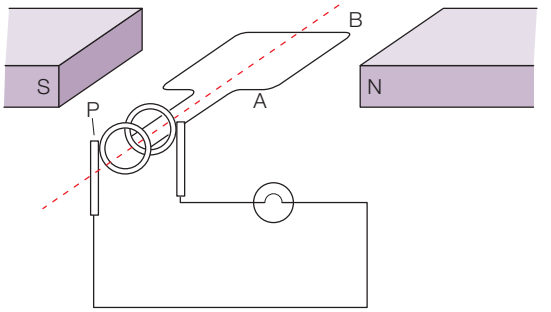
(v) Briefly describe *one* practical use of the device shown.

(vi) Name *one* other device that works on the same principle as the device shown.

5 A voltage of 110V is applied to the primary coil of a transformer. If the primary coil has 2000 turns and the secondary coil has 500 turns, calculate the output voltage of the transformer.

6 A transformer is used with the 240V a.c. supply to deliver 9.0A at 8.0V to a heating coil. If 10% of the energy taken from the supply is dissipated in the transformer itself, what is the current in the primary winding?

7 The diagram shows a coil that is being rotated in the magnetic field in a certain device. At time $t = 0$, the segment AB is moving downwards.



(i) On the diagram, indicate the direction of the current through AB at time $t = 0$.

(ii) Sketch a graph showing the variation in current (y-axis) vs. time (x-axis) for one complete cycle of the oil's rotation, starting at time $t = 0$.

(Mark the time axis in quarters of a cycle.)

(iii) What is the name given to P?

(iv) What is the name given to the process taking place in the rotating coil?

(v) State *one* use of the device shown in the diagram.

8 Describe *three* features of a practical transformer that contribute to its efficiency.

9 In an electric guitar, when a steel guitar string vibrates next to a device called a 'pick-up', a varying e.m.f. is obtained from the pick-up. What does this suggest about the construction of the pick-up?

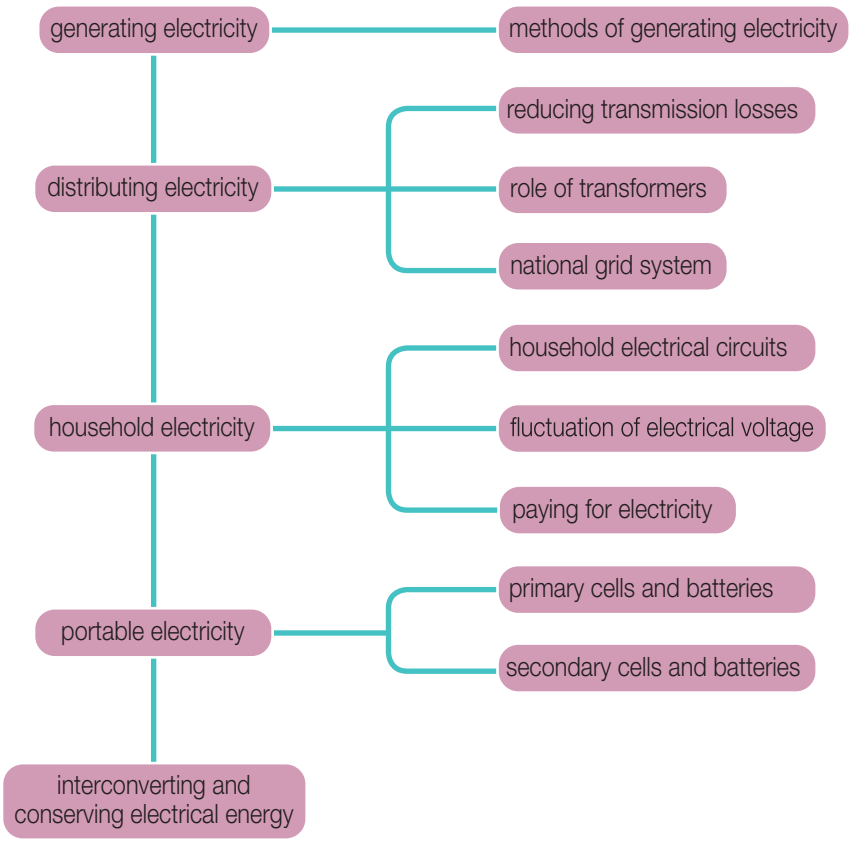
29 Electricity for Home and Society

By the end of this chapter you should be able to:

- describe the main methods of electricity generation in the Caribbean
- describe the distribution of electrical energy from the generating station to the consumer
- describe electrical circuits used in the home, and discuss reasons for using series, parallel and ring connections
- state the international insulation colour code used on wires carrying electricity in the home
- relate electrical safety to correct wire sizes, suitable fuses or circuit breakers and earthing
- state the adverse effects of connecting electrical appliances to incorrect or fluctuating voltage supplies
- perform calculations on the cost of electrical energy expressed in kilowatt-hours (kWh)
- draw a diagram of a dry cell and explain the functions of its various parts;
- state the difference between primary and secondary cells, and their relative advantages and disadvantages
- draw a circuit diagram to show how a secondary cell can be recharged
- identify conversions of electrical energy to other forms and vice versa
- discuss the importance of conserving electrical energy, and means of doing so



Figure 29.1 Inside a modern supermarket.



Introduction: electricity at work

ITQ1

How many items in figure 29.1 can you see that use electricity?

Energy is needed to operate equipment and appliances. The most widely used form of energy today is **electrical energy**. Electrical energy can be sent at very high speeds (almost at the speed of light) along wires to local and distant places. Electricity is also easily converted into other forms of energy, and other forms of energy can easily be converted to electrical energy.

Generating electricity

Fossil fuel methods

fossil fuel

In most Caribbean counties electrical power is generated from **fossil fuels** (figure 29.2). These fuels were formed by a very slow process over millions of years as plant remains were transformed under pressure into crude oil, coal or natural gas. In the Caribbean, crude oil and natural gas are found in abundance in Trinidad and Tobago. Natural gas and liquid fractions, like bunker C fuel, distilled from crude oil, are used in electricity generation plants.



Figure 29.2 Fossil fuel power. This power plant in the heart of Port of Spain, Trinidad and Tobago, works quietly burning natural gas.

heat energy furnace

turbine

Figure 29.3 shows how electricity is generated at one type of fossil fuel power station. Oil or natural gas is burnt in a furnace to produce **heat energy**. The heat produced by the **furnace** converts water in the tubes of a boiler into steam, which leaves the boiler at very high pressure. The high-pressure steam pushes against the blades of a **turbine**, causing them to spin.

condenser

The steam gives up much of its heat energy to the turbine and leaves the turbine at low pressure. The low-pressure steam is cooled by a **condenser** and converted back to water. The water is then pumped back to the boiler, where it is converted into steam at high pressure, and the whole process is repeated as before.

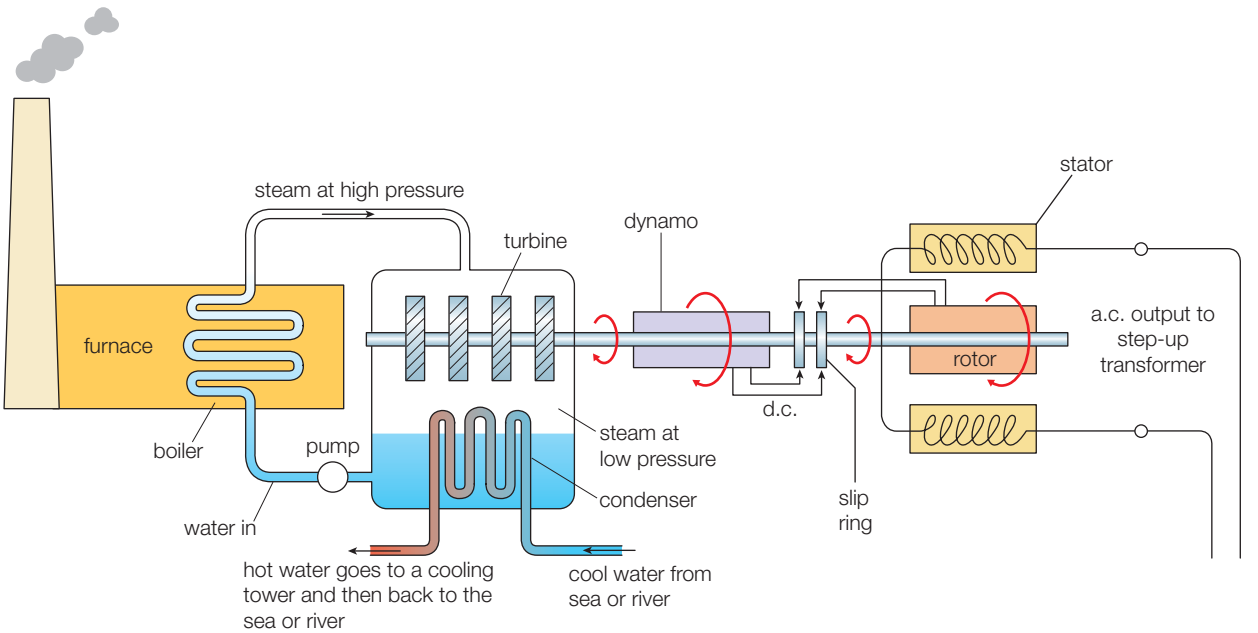


Figure 29.3 Generating electricity by burning fossil fuel.

The cool water entering the condenser tubes is usually taken from a nearby river or the sea. While in the condenser tubes it becomes quite hot as it absorbs the latent heat from the low-pressure steam as it cools down to form water.

dynamo ›

As the turbine spins, a small **dynamo** attached to a shaft turns and generates a direct current. This direct current is fed to a rotating coil, called a rotor, by means of slip-rings attached to the coil. Since the rotor has a core of soft iron, it then becomes a powerful **electromagnet** (see figure 28.23, page 441).

electromagnet ›

The rotating electromagnet induces an alternating e.m.f. in a stationary coil, called the stator, that surrounds it. The very large output voltage produced across the stator is then applied directly to a transformer, where the voltage is stepped up.

internal combustion engine ›

Other types of fossil fuel power plants use **internal combustion engines**, such as diesel engines, instead of furnaces and boilers, to obtain rotational kinetic energy from the energy stored in the fossil fuels. These engines are noisier and need much more maintenance. The rotating shaft is attached to a dynamo and rotor to induce a current in a stator, as described above.

Other methods of electrical power generation

wind ›

In Antigua the power of the **wind** has been put to use. Figure 29.4 shows a 120 kW Darrieus rotor, wind-powered, vertical-axis generator. The turbine looks somewhat like a giant egg-beater. The use of wind-powered generators of this capacity is feasible on exposed sites where there is a fairly steady wind of over 6 m s^{-1}. Sites like these can be found near the ocean. Jamaica and Barbados have also been introducing 'wind farms' which convert wind energy to electrical energy (figure 29.5). However, there are pros and cons involved in such developments. For example, although there is hardly any chemical pollution resulting from the energy conversion, wind turbines are somewhat noisy and unsightly – the latter being of concern to scenic tourist destinations.

Internet search term: wind energy in the Caribbean



Figure 29.4 A 120 kW Darrieus rotor, wind-powered generator that has been tried in Antigua.



Figure 29.5 Wind farm at Manchester, Jamaica. Barbados is also considering developments such as these.



Figure 29.6 Solar cells can be used for power in a difficult environment.

Solar power is used in Guyana to provide electricity to a small hospital at Orealla, an Amerindian settlement in Guyana's hinterland. During the day, solar-voltaic cells convert light energy from the Sun into electrical energy. This energy is stored in large batteries and used as needed by the hospital, mainly for refrigeration of medical supplies and fluorescent lighting. Solar power plants have no moving parts, and so they require very little maintenance. This makes them very useful in situations where technical support is limited, such as on the Moon and Mars (figure 29.6).

hydroelectricity ›

IT Q2

In which of the States associated with CARICOM is the massive hydroelectric power plant shown in figure 29.7?



Figure 29.7 A large hydroelectric power plant.

Hydroelectric power plants make use of gravitational potential energy stored by water in a high region (figure 29.7). As the water falls to lower regions, the potential energy is changed into the kinetic energy of moving water. The moving water causes a turbine to spin. The turbine is connected by a shaft to a generator that converts the rotational kinetic energy of the turbine into electrical energy.

Guyana has large rivers flowing from its mountainous regions. However, there are only a few small-scale hydroelectric power plants in the hinterland. These plants make use of the kinetic energy of naturally occurring waterfalls.

After the initial costs of starting up, hydroelectric power is cheap, since fuel is not needed. However, for large-scale power generation, rivers must be dammed to create reservoirs, which must not become depleted during dry seasons. The construction of such reservoirs involves severe ecological disturbance, as trees would have to be cut down and animals displaced from their habitats. Human populations may also have to be moved. The relatively still reservoirs themselves might even serve as a breeding ground for mosquitoes, carriers of the dreaded malaria and dengue diseases. Hence, any large hydroelectric scheme must be carefully planned. An ecological benefit from a successful hydroelectricity scheme is that generation of electricity by this method leads to very little pollution.



Figure 29.8 The electric eel, Electrophorus electric.

Bioelectricity! The electric eel, *Electrophorus electric* (figure 29.8), found in South America, can generate as much as 600 volts – enough to stun a person.

BIOLOGY: ecology

Distributing electricity

Electrical energy is most efficiently generated and distributed using alternating current (see page 446). Most Caribbean countries, and also the USA, utilise an a.c. frequency of 60 Hz. A few territories use a frequency of 50 Hz.

Minimising power losses

When current travels along a wire, the wire becomes hot and heat energy is lost to the atmosphere. Power is lost according to the formula

$$P = I^2R$$

where P is the power lost, I is the current in the wire and R is the resistance of the wire (page 412). In order to minimise power loss, two approaches are used in distribution systems.

One is to keep the resistance (R) of the conducting wires as small as possible. The transmission wires therefore have larger diameters, since thick wires have less resistance per unit length than thin wires.

A second approach is to use small currents in the wires. To achieve this, the power is sent at very high voltages. In Worked example 28.4, we saw that a power of 1 000 000 W, sent at 10 000 V, results in a power loss of 20 000 W along the lines. The same power, sent at 100 000 V along the same lines, results in a power loss of only 200 W!

IT Q3

How does water get to high regions, resulting in the storage of gravitational potential energy?

IT Q4

100 000 W of power is distributed along wires of total resistance 2 W. How much power would be lost if the 100 000 W were sent using: (i) 1000 V; (ii) 10 000 V? (Hint: look at Worked example 28.4, page 445.)

The role of transformers

Transformers are used to raise or lower voltages as required along the distribution lines from the generator to the user.

Figure 29.9 shows a simplified diagram of a power distribution system involving a single generator. Note the use of the transformer to step up the voltage (and hence step down the current) from the generator.

Transformers at substations along the distribution lines, and between the substations and homes and factories, step down the high voltages to the values required by users.

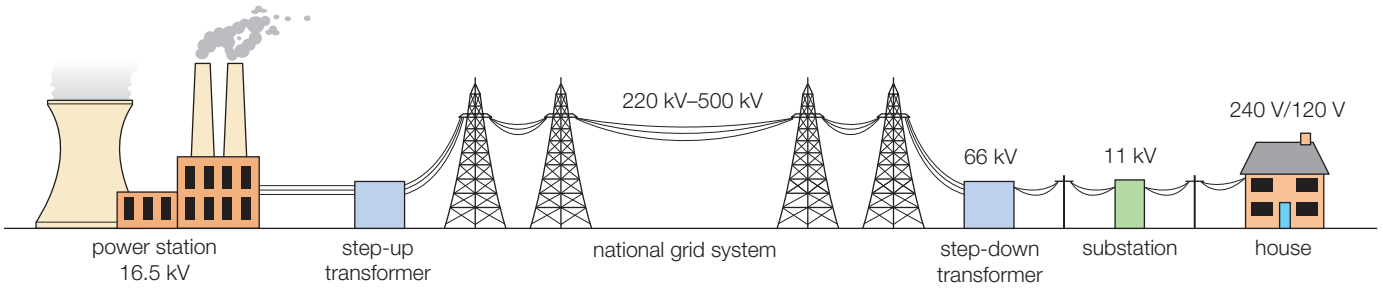


Figure 29.9 Power distribution from a generating station.

National grid system

In many territories, the outputs of several generators are connected in parallel to form a national grid system. One advantage of the grid system is that several generators can be linked into one system to provide enough power when the demand for electricity is high.

ITQ5

Can you think of another advantage of the grid system?

Household electric circuits

Parallel wiring

Parallel wiring is used in household circuits rather than series connections, so that appliances can be controlled individually without affecting other devices. (See chapter 26, pages 400 and 403 for a discussion of series and parallel circuits.)

If any one switch in a series circuit (figure 29.10) is open, there will be no current in the circuit. Further, if any one of the devices in a series circuit 'blows', and creates an open circuit, again, there will be no current and the other appliances in the circuit will not work. For this reason, series connections are only used in special circuits, for example fairy lights, where lamps are connected in series so that one switch can turn off and on an entire set of lamps (figure 29.11).

ITQ6

Which lamp(s) in figure 29.10 should light when: (i) switches 1 and 2 are closed; (ii) switch 1 is open and switch 2 is closed; and (iii) switch 1 is closed and switch 2 is open?

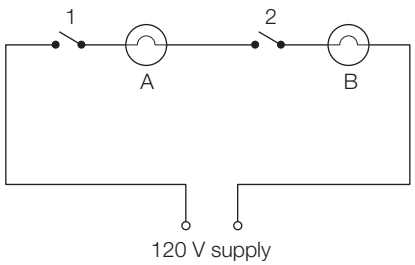


Figure 29.10 Two lamps connected in series.

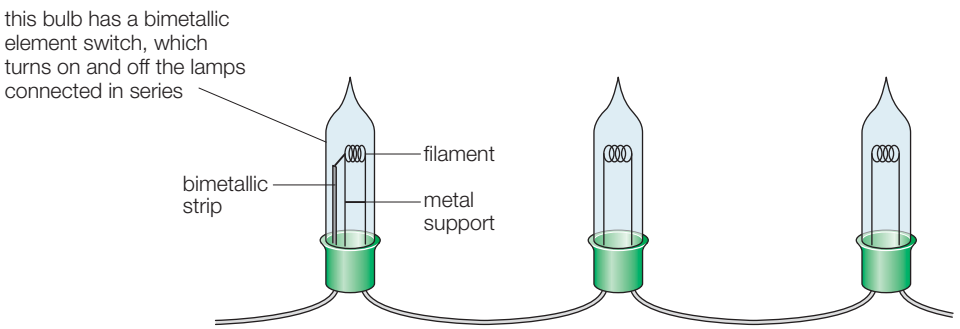


Figure 29.11 Fairy lights connected in series.

IT·Q·7

Which lamp(s) in figure 29.12 should light when: (i) switches 1 and 2 are closed; (ii) switch 1 is open and switch 2 is closed, and (iii) switch 1 is closed and switch 2 is open?

Figure 29.12 shows two lamps connected in parallel. In the circuit shown, as long as any one switch is closed, there will be a closed path for current to pass through that switch to the appliance to which it is connected. In a parallel arrangement, therefore, each appliance can be controlled independently, without affecting the operation of other devices. Further, whereas in a series circuit the components share the supply voltage in proportion to their resistance (see Worked example 26.2, page 402), in a parallel circuit, all components receive the supply voltage and are thus able to function properly.

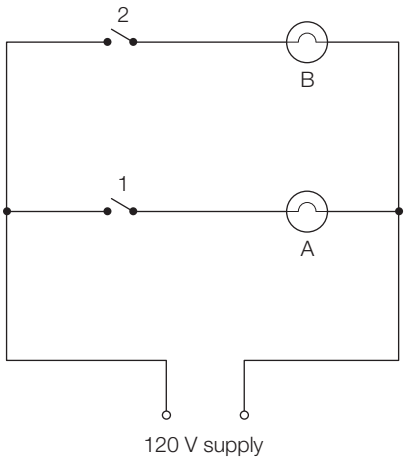


Figure 29.12 Two lamps connected in parallel.

Electrical wiring in a house

Household circuit wiring practices may vary from territory to territory, depending on the wiring codes used in each. The wiring may also vary depending on whether the dwelling is a separate house or part of a set of apartments.

Live (L), earth (E) and neutral (N) wires

neutral ›

live ›

Two wires enter the house from the electricity company's supply. One wire is earthed at the local substation and is called the **neutral**. Because this wire is connected to earth, its voltage is the Earth's voltage, that is, zero. You are not likely to get shocked if you accidentally touch this wire. The other wire is called the **live** wire. The voltage on this wire is alternately positive and negative. You can be shocked if you accidentally touch this wire, since it is at a high voltage almost all the time.

Colour-coding and plugs

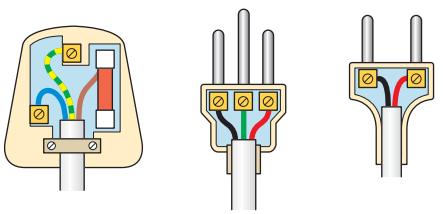


Figure 29.13 Colour-coding of wires attached to 3-pin plugs and a 2-pin plug.

The plastic insulation on electrical wires used in household wiring or attached to plugs or equipment is often colour-coded. In the international system, the live wire is coloured brown, the neutral is blue, and the earth is yellow and green (figure 29.13). Other systems are in use in various countries. In one such system, the live wire is coloured red, the neutral is black, and the earth is green. Two-pin plugs connected to electrical equipment often have one broad and one narrow blade that can fit only one way into an outlet. This ensures that the live wire is connected to the switch side of the device being used.

Earthing

Outlets usually carry a third wire, the earth wire, whose insulation is coloured yellow/green (figure 29.13). This wire is connected to earth outside the home, usually by means of a metal pole driven into the ground. The earth wire is connected electrically to the casing (called the 'chassis') of electrical devices such as refrigerators, cookers or computers.

Should there be a fault so that the live wire in an appliance touches the casing, a short-circuit to earth would be created; there would be a large current from the live wire via the metal casing to the earth wire. Without an earth wire, the appliance casing would be raised to a high voltage; it would be 'live', and a person touching the case could receive a very serious electric shock,

since current would pass to earth through their body. However, as the earth wire takes almost all the current to earth, very little passes through the body of anyone touching an earthed case (figure 29.14).

If there is no fuse or circuit-breaker in the circuit, and the large current flows to earth for any length of time, the wiring would burn, with disastrous consequences. However, a fuse or circuit-breaker present in the circuit would blow or trip, cutting off the current.

Note that it is extremely important to have a suitable fuse in any circuit that has an earth wire, since it possible that such a circuit might at some time have to conduct a very large current. It is also very important that the earth wire be connected firmly to the earthing pole at all times so that current can flow readily to the earth if the metal case of an appliance happens to touch a live wire.



Figure 29.14 Suppose a faulty live wire inside the cooker is touching the metal casing. If the cooker is correctly connected to the power supply and earth, anyone touching the cooker casing will not get a shock.

ITQ8

In figure 29.14, why is the current through the householder's body almost zero and that through the earth wire so large, if the metal casing is properly connected to earth?

The entrance unit

main fuse ›
entrance unit ›

The live wire is connected to the electricity company's **main fuse** inside an **entrance unit** (figure 29.15). If the current levels in the house become dangerously high, this fuse 'blows' and opens the circuit, so that none of the wires in the house circuit are then live.

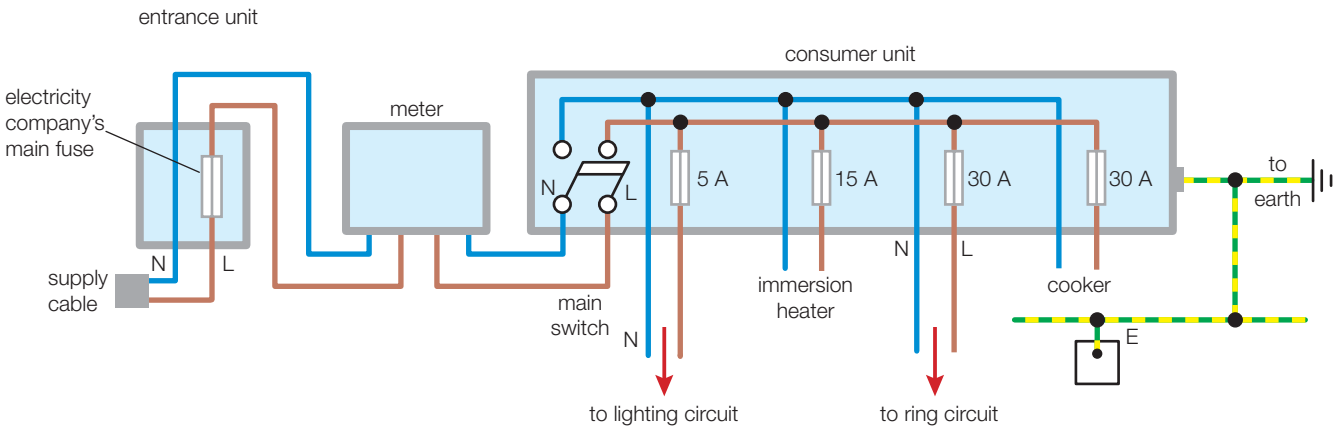


Figure 29.15 Entrance unit, meter and consumer unit.

meter ›

The incoming current passes through the **meter**, which records the amount of electrical energy converted within the house.

The consumer unit

consumer unit distribution box ›

The meter is connected to a **consumer unit** (also called a **distribution box**) via a double-pole single-throw (DPST) main switch (figure 29.15). When 'off', this switch completely isolates the house circuit from any electrical contact with the live or neutral wire of the power supply. (In some countries, the main switch and the meter are placed next to each other outside the house.) Because electricity can be so dangerous, it is wise to know where this main switch is situated so that power can be shut off in an emergency.

ITQ9

What are the items labelled A and B in figure 29.16?

Inside the consumer unit are fuses or circuit-breakers (figure 29.16) to protect various circuits in the house. Nowadays, rather than fuses, increasing use is made of circuit-breakers, which 'trip' and open a circuit that is drawing too much current. Electromagnetic circuit-breakers are now preferred because they act faster than fuses and also can be reset easily. If fuses are used, the consumer unit is called a 'fuse-box'; it is called a 'contact-breaker panel' if circuit-breakers are used.

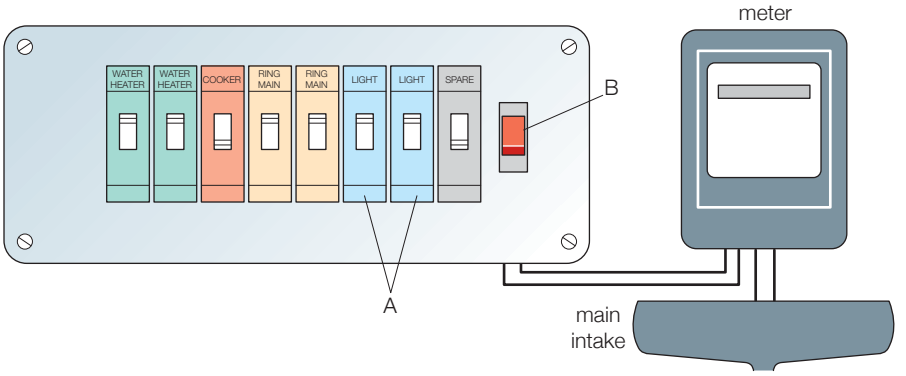


Figure 29.16 A consumer unit, main switch, electricity meter and entrance unit. (Note: In some territories, only the main switch, along with the meter, is located outside the house. The consumer unit is placed inside.)

Parallel, ring and high-power circuits

Figure 29.17 shows the physical layout of one type of electrical wiring in a house.

Two circuit systems are generally used; a simple parallel circuit and a ring parallel circuit. Lamps (which use small currents) are usually wired in a simple parallel arrangement. Electrical outlets are also connected in parallel but they are usually wired on the ring circuit. These outlets are suitable for devices that use moderate power. Appliances with a very high power requirement, such as stoves, are usually connected in parallel directly to the main power distribution box.

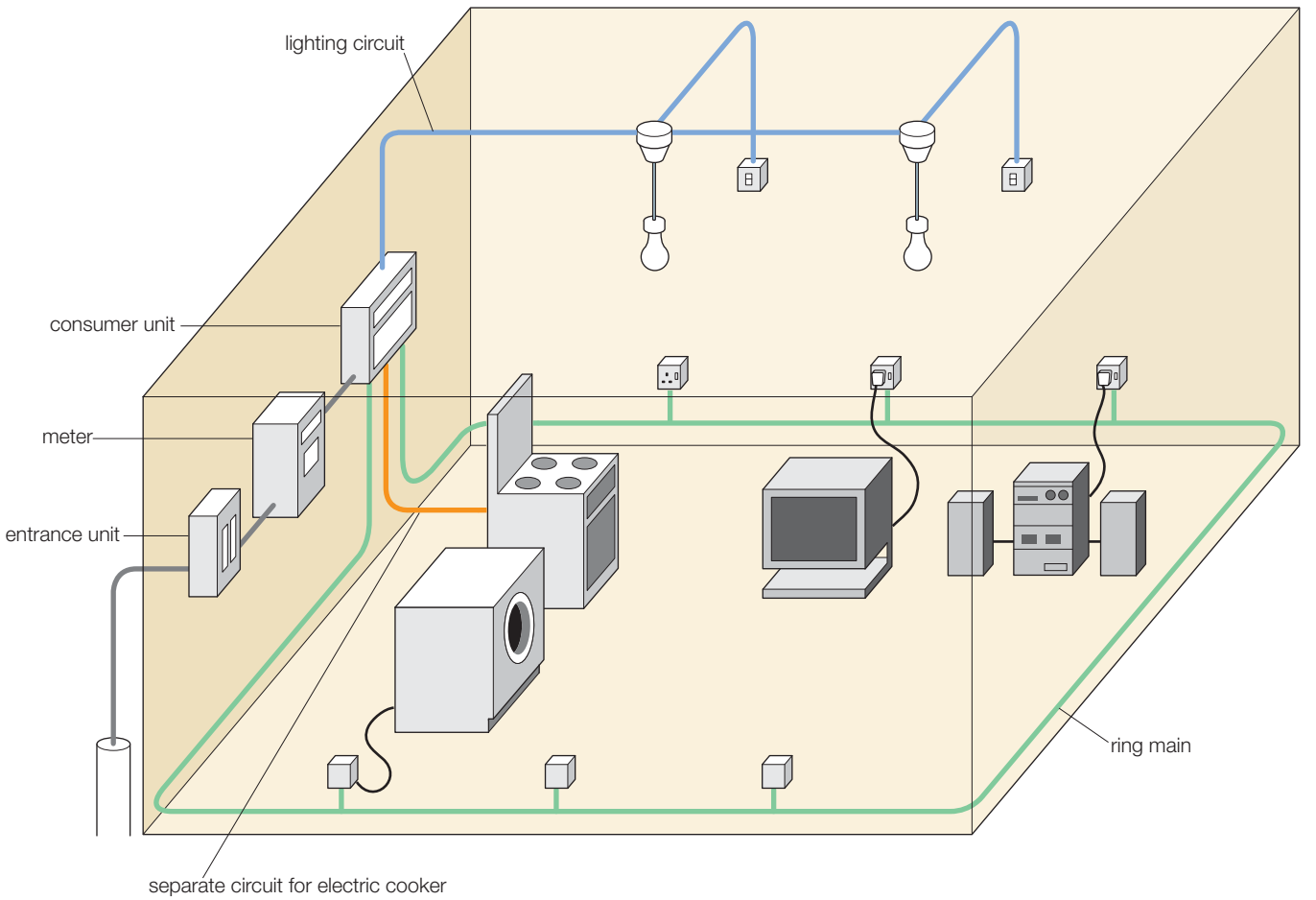


Figure 29.17 Physical layout of electrical wiring in a house showing lighting, ring and high power circuits. Inside each of the orange and green cables are the live, earth and neutral wires. Inside the blue cable, only live and neutral wires are enclosed.

Table 29.1 lists some household equipment, with nominal power ratings and the current that would be drawn by the equipment when operated at 120 V and 240 V.

Table 29.1 Electrical requirements of selected household equipment.

Equipment	Nominal power/W	Current in amperes when operating at 120 V	Current in amperes when operating at 240 V
radio	10	0.083	0.042
television	100	0.833	0.417
lamp	60	0.50	0.25
hair dryer	1250	10.4	5.2
iron	500	4.17	2.08
stove	5000	41.7	20.8

On account of the very high currents involved at lower operating voltages, high-power appliances are usually designed for 240 V rather than 120 V operation. Some households therefore carry both a 120 V and a 240 V supply. The high-power devices, designed for 240 V operation, are connected to the 240 V supply and use currents that are not too large (see Worked example 29.1). The cost of wiring such houses is reduced, since there is less need for very thick wires, which are quite expensive but will carry large currents safely.

The mains voltages across the Caribbean are not the same. Some Caribbean territories use 110/220V, 115/230 V and others 120/240 V.

Worked example 29.1: 110 V or 220 V?

Some of the household electrical circuits in the Caribbean utilise a supply voltage of 110 V or 220 V or both. From the point of view of power loss along wires, which of the two voltages is the better option?

Solution

To answer this, suppose 440 W of power is used: (i) by what factor is current increased by using a 110 V rather than a 220 V supply, and (ii) by what factor of power lost by using a 110 V rather than a 220 V supply?

(i) As power (P) = current (I) × voltage (V), then $I = P/V$. Therefore,

$$\text{current in 110 V circuit} = \frac{440\,\text{W}}{110\,\text{V}} = 4\,\text{A}$$

$$\text{current in 220 V circuit} = \frac{440\,\text{W}}{220\,\text{V}} = 2\,\text{A}$$

Hence the current is doubled, that is increased by a factor of 2, in the 110 V circuit.

(ii) Since power loss is directly proportional to the square of the current in a circuit ($P = I^2R$), then power loss in the 110 V circuit will be increased by a factor of 2^2, i.e. 4.

From the point of view of power loss, it is therefore more economical to use a 220 V power supply than a 110 V supply. However, the higher the voltage of the supply, the more hazards are encountered during usage; hence more safety features are inserted into the wiring of such a supply, resulting in greater initial expense.

Safety warning! You should **never** touch any exposed conductor of a mains circuit – not even the neutral wire! Only qualified electrical personnel should handle mains circuits.



Figure 29.18 Electrical hazard warning symbol.

Figure 29.18 shows a warning symbol for electrical hazard.

Lighting circuits

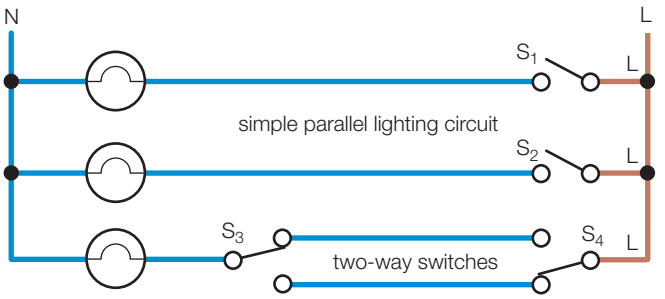


Figure 29.19 Wiring of lights in parallel in a household circuit.

Lights are wired in simple parallel arrangements (figure 29.19). Switches are placed on the live lines (L) leading to the lights. When a switch is turned off, there will be no live connection to a lamp, so the lamp can be safely handled. The switches S_1 and S_2 are single-pole single-throw (SPST) switches. S_3 and S_4 are single-pole double-throw (SPDT) switches, which are wired in such a way that the same lamp can be switched on or off from two different places, and these switches are sometimes called 'two-way' switches. Staircase lamps are usually wired using these kinds of switches, with one placed at the bottom of the stairs and the other at the top.

When there is excessive current in a circuit, a fuse 'blows' (or a circuit-breaker 'trips') so that there is no current, thus preventing overheating. The choice of appropriate fuses and wires is discussed in the section 'Short-circuits and overloads' in chapter 25 (page 394).

Household circuit-breakers come in standard values, such as 5 A, 10 A, 15 A, 20 A, 30 A and 40 A. Fuses in appliance plugs are usually rated at 1 A, 2 A, 5 A or 13 A. Cartridge fuses are usually used in these plugs.

ITQ10

Using the information about fuse ratings given above and in table 29.1, what fuse rating would be suitable for the plug of a 110 V hair dryer?

Worked example 29.2: Number of lamps in a circuit

How many 100 W lamps can be accommodated on a 120 V circuit that is protected by a 10 A circuit-breaker?

Solution

Power (P) = current (I) × voltage (V). So total power accommodated for a 10 A circuit breaker is given by

$$P = 10\,\text{A} \times 120\,\text{V}$$

$$= 1200\,\text{W}$$

Hence, the maximum number of lamps = $\frac{1200}{100} = 12$.

(However, allowing for supply voltage fluctuations, 11 or 10 lamps might be a better choice.)

High-power circuits

High-power devices such as water heaters and stoves are usually wired directly from the supply in the consumer unit. Fuses (or circuit-breakers) with large ampere ratings are used so as to allow large currents to flow to these devices. The wires connecting these devices to the mains are thick so that they have very little resistance, and there is little heating as current flows through.

Worked example 29.3: Electric stove

What current rating of circuit-breaker would be appropriate for a 220 V, 5 kW cooker?

Solution

Power (P) = current (I) × voltage (V). Hence

$$I = \frac{I}{V}$$

$$= \frac{5000\,\text{W}}{200\,\text{V}}$$

$$\approx 21\,\text{A}$$

A 30 A breaker would allow for fluctuations in the current while not permitting excessive current (such as would occur in a short-circuit).

Ring circuits

Electrical outlets are usually wired on a ring circuit (figure 29.20). The ring circuit is also a parallel circuit but, unlike the lighting circuit (figure 29.20 (a)), it allows current to flow to an outlet via two routes (figure 29.20 (b)). Thus, each live wire in the ring carries only half the current required by the devices. The ring circuit can therefore safely have more outlets than would be possible with the simple parallel arrangement used for lighting circuits.

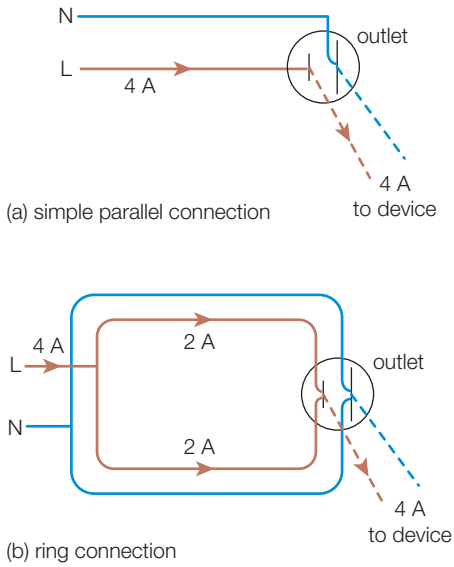


Figure 29.20 How a ring circuit helps to reduce the current in a circuit wire. In the simple parallel connection, the current reaches the outlet via only one live wire.

Surges, spikes and low voltages

Electrical and electronic equipment are often ruined by fluctuations in voltage (figure 29.21). The damages are caused by surges, **spikes** or low voltages.

spike ›
surge ›

A **surge** is an increase in voltage that is significantly above the normal level and lasts for more than 3 nanoseconds (3×10^{-9} s). Surges occur when high-powered devices (i.e. devices taking large currents), such as air-conditioners, refrigerators and motorised equipment, are switched on or off. A surge can also occur when a generator is switched on to or off from a grid system.

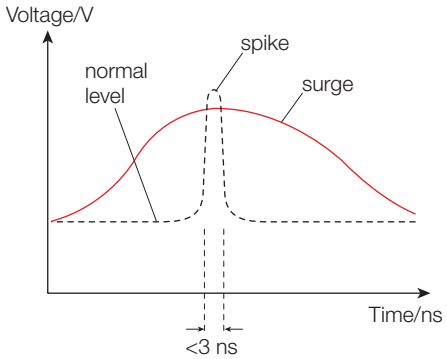


Figure 29.21 Surge and spike voltage variations with time.

A spike is also an increase in voltage significantly above the normal level that lasts, however, for less than 3 nanoseconds. The brief but massive discharges that take place during lightning tend to induce spike voltages in power lines.

Both surges and spikes inflict damage especially on electronic equipment such as computers and amplifiers. Electronic components such as those present in semiconductor devices can easily be destroyed by high voltages (see chapter 30 for a discussion of semiconductor devices). The resistive elements in appliances such as irons and lamps will overheat, and may even burn up since the large voltages result in large currents. Wires connected to these devices can also overheat on account of the excessive current.

A surge-and-spike protector prevents the voltage, and hence the current that reaches the equipment, from rising to dangerously high values.

low voltages ›

Low voltages occur when circuits are overloaded. Some voltage drop occurs along feeder lines because of the large current in the wires. Low-voltage damage occurs usually to motors (e.g. refrigerator motors). There is not enough current to turn the motor fast enough, or at all, and hence the low current passing through the motor produces heat in the motor coils instead of turning the motor and doing mechanical work. It is this heat which eventually damages the coils of the motors.

Surge and spike protection

According to one manufacturer, research shows that only about 5% of damage to equipment results from high-voltage (greater than 200 V) surges and spikes. Most damage is done by low-voltage surges and spikes. Metal oxide varistor (MOV) protectors are effective only for high-voltage surges and spikes, such as may occur during a lightning storm. Transformer-based filtering (TBF) offers efficient protection from low-voltage surges and spikes.

Paying for electricity

We pay for electrical energy converted by appliances in our homes. (Note that it is not current that we pay for. We do not consume current, but only use it as a vehicle to bring electrical energy to our homes and factories. We might say that having done its job, the current returns to where it came from – the power station.)

We calculate energy expended (see page 412) using the formula

energy = power × time

kilowatt-hour (kWh) ›

Most power companies measure power in kilowatts (kW) and time of usage in hours (h). Hence power companies measure energy in kW × h or **kWh** (**kilowatt-hours**). A kilowatt-hour is called the 'unit' of electricity used. One kilowatt-hour of energy is used when a 1 kW device converts energy for 1 hour. For example, a 3000 W stove, operating for 4 hours converts 3 kW × 4 h, or 12 units (12 kWh), of electrical energy.

When an electricity meter is read, the number of units used is found by subtracting the previous reading from the present reading. An electricity company may charge a fixed rate for each unit used and an additional 'fuel adjustment costs' rate, which depends on the prevailing cost of the fuel being used at the power station.

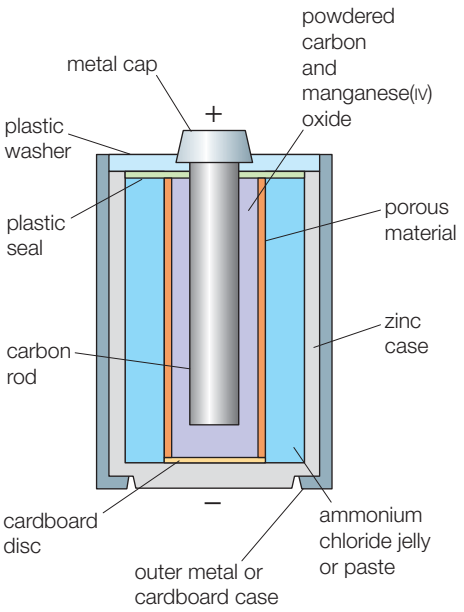


Figure 29.22 A modern Leclanché dry cell.

Worked example 29.4: Cost of electricity

A company's fixed rate is 17 ¢ per unit, and the fuel adjustment rate is 2.3 ¢ per unit. A consumer uses a 200 W floodlight for 36 hours (e.g. 4 hours every night for nine nights). How much would the energy used cost the consumer?

Solution

The energy used = 200 W × 36 h = 7200 Wh = 7.2 kWh

The subtotal for the fixed rate = 7.2 × 17¢ = $1.22

The fuel adjustment charge = 7.2 × 2.3¢ = 17¢

The total charge = $1.22 + $0.17 = $1.39.

Portable electricity

In chapter 25, we introduced Volta's simple, yet important, invention – the electric battery. Volta's battery was made up of 'cells', each consisting of two different metals in contact with a conducting solution (an electrolyte). Modern cells have made it possible for electricity to be available practically anywhere.

Figure 29.22 shows a modern cell, which makes use of an electrolyte in the form of a thick paste. It is called a 'dry' cell, since the electrolyte is not in a liquid state.

ITQ11

Name one advantage of using a paste electrolyte rather than a liquid one.

Carbon-salt-zinc primary cells and batteries

primary cell ›
Leclanché dry cell ›

A **primary cell** cannot be recharged (by applying a reverse voltage) after it has run down. The **Leclanché dry cell** shown in figure 29.22 is perhaps the cheapest and most popular primary cell. This cell has a carbon rod and a zinc case as its two electrodes. The electrolyte is moist ammonium chloride (a salt) in a jelly-like paste. The atoms of zinc in contact with the moist ammonium chloride lose electrons and become zinc ions:

CHEMISTRY: electro-chemistry

$$Zn \rightarrow Zn^{2+} + 2e^-$$

When a wire connects the electrodes, electrons liberated from the zinc atoms flow to the carbon rod and along the wire. Hydrogen ions from moisture in the ammonium chloride paste migrate across the porous material to the carbon rod and accept the electrons to form hydrogen atoms. Hydrogen atoms then combine to form hydrogen molecules:

$$H^+ + e^- \rightarrow H \text{ (atom)}$$

The electrodes of a cell are the two conductors that are in contact with the electrolyte. When electrodes are connected by a wire, electrons flow from one electrode (called the negative electrode) to the other.

polarisation ›

depolariser ›

$H + H \rightarrow H_2$ (gas)

The layer of hydrogen gas formed around the carbon partly insulates the carbon electrode, restricting the flow of electrons to the carbon rod. Further, the layer of hydrogen gas also functions as an electrode, which produces an e.m.f. opposed to that of the cell. This is called **polarisation**. The insulation and polarisation effects could result in the cell ceasing to function. Powdered manganese(IV) oxide, an oxidising agent placed around the carbon rod, reacts with the hydrogen as soon as it is formed and removes it. The manganese(IV) oxide is called a **depolariser**. The cell will then deliver current until all the depolariser is used up.

Typically, an e.m.f. of approximately 1.5 V is found across the carbon and zinc electrodes of a freshly made Leclanché cell. When stacked or connected in series, the cells form batteries, such as the popular 6 V and 9 V batteries.

Leclanché dry cells work well where moderate currents (e.g. up to about 1 A) are needed intermittently, for instance in flashlights. When large currents are drawn for sustained periods, the cells polarise rapidly and become 'dead'. The main problem with the Leclanché cell, however, is that once the chemicals are used up, the cell cannot be recharged.

Leclanché cells come in various sizes, that is, AAA, AA, C and D. Each of the cells has the same e.m.f. since they use the same chemicals as electrodes and electrolytes. The larger the size of the cell, however, the more chemicals the cell contains. This enables the cell to deliver current longer under normal operating conditions. The length of time a cell an deliver current in normal use is a measure of what manufacturers call the 'life' of the cell. A cell or battery's 'shelf life' refers to the time a cell can remain unused and still function when needed (have another look at Practical activity 25.2, 'Making and investigating battery cells').

ITQ12

The manganese(IV) oxide is mixed with powdered carbon (figure 29.22). Why is this done?

ITQ13

Some people claim that putting an almost 'dead' cell in sunlight, in a freezer, or tapping it, causes it to be 'recharged'. Explain why these might work.

ITQ14

How many 1.5 V cells are present in a 9 V battery?

Other types of primary cells and batteries

alkaline cell ›

Alkaline primary cells use potassium hydroxide rather than ammonium chloride as their electrolyte. The positive electrode (called the anode) is powdered carbon. The negative electrode (called the cathode) is powdered zinc. Compared with Leclanché cells, **alkaline cells** can deliver greater current, have longer life and shelf life, and maintain a more steady e.m.f. as they discharge.

There are other types of primary cells. Mercury oxide and silver oxide cells provide very stable e.m.f.s, and so they are useful in electronic watches where a stable voltage, rather than a high current, is required. These cells, however, are expensive. Lithium cells have very long life, as well as a very long shelf life of 5 years or more, and they are used as back-up cells for computer memories.

Mercury oxide batteries: Mercury compounds are toxic, and disposal of mercury oxide batteries after use poses an environmental problem.

Internet search terms: primary cells and batteries

Secondary cells and batteries

secondary cell ›

Like primary cells, a **secondary cell** consists of two dissimilar electrodes in contact with an electrolyte, but a secondary cell can be recharged by connecting a reverse voltage across its electrodes. The most common secondary cell is the lead-acid cell (figure 29.23) used in motor vehicles. Most cars have a 12 volt battery consisting of six 2 volt '**lead-acid**' cells.

lead-acid cell ›

The positive electrode in a lead-acid cell consists of plates of lead dioxide.

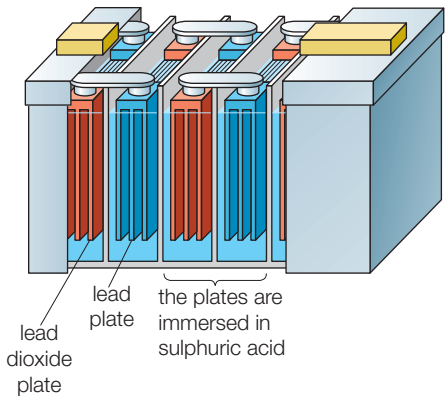


Figure 29.23 Inside a lead-acid battery.

The negative electrode is made of plates of 'spongy' lead. (The 'spongy' texture increases the surface area of the lead in contact with the electrolyte.) The electrolyte is dilute sulphuric acid with a density of about 1.25 g cm^{-3} (which is strong enough to char clothing and burn skin). The e.m.f. of a lead-acid cell, ready for use, is approximately 2.2 V.

During discharge, the lead dioxide and lead react with the sulphuric acid and are converted into lead sulphate. Energy taken from the cell during this process enables electrons to flow from the negative electrode (the lead plates) to the external devices that are connected to the battery. The e.m.f. of the cell falls during discharge to about 1.8 V. The density of the electrolyte drops eventually to about 1.15 g cm^{-3}, resulting in a 'flat' battery.

A 12 V lead-acid battery can be recharged by connecting a d.c. source of about 15 V in reverse across the electrodes (figure 29.24). The positive of the charging voltage is connected to the positive of the battery. The chemical reactions are reversed, and the density of the sulphuric acid rises to its original value as the acid is re-formed.

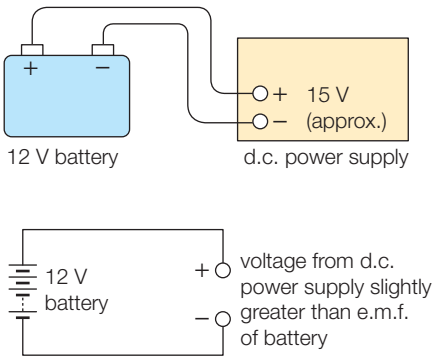


Figure 29.24 Recharging circuit for a secondary battery.

Some advantages of lead-acid batteries are as follows:

- They can store a very much larger charge than primary cells of the same size and so can deliver current for a much longer time.
- They have a low internal resistance and thus can deliver very large currents. (Vehicle starter motors require very large currents, as much as 100 A and greater!)
- They can be easily recharged.

Some of their disadvantages are as follows:

- The older type have a very short shelf life, as chemical reactions taking place between the electrodes and the electrolyte produce a hard, unreactive type of lead sulphate when the batteries lie unused for some time.
- They are very heavy.
- They must be handled very carefully to prevent spillage of sulphuric acid, especially on to skin.
- An accidental short-circuit across the electrodes can produce a very large current that can be enough to overheat a battery and even result in an explosion. Drawing a large current from the cell would almost certainly damage it.
- After a battery is fully charged, water in the cells becomes electrolysed to produce hydrogen and oxygen; a spark in contact with this mixture can produce an explosion.
- Safe disposal after use can pose a problem, since lead is *toxic* to the environment.
- They require careful maintenance: the cells must be topped up with pure water so that the plates are just covered; density checks of the electrolytes must be done, and the battery must have continuous use to reduce the formation of the hard lead sulphate compound.

Modern lead-acid batteries have a long shelf life and are almost maintenance-free.

nickel-cadmium (nicad) cell ›

Nickel-cadmium (nicad) cells are another type of secondary (rechargeable) cell. The negative electrode is made of cadmium and the positive one of nickel. The electrolyte is a strong solution of potassium hydroxide. The e.m.f. of a fully charged nicad cell is about 1.2 V. Nicad cells can deliver large currents. However, in practice, the number of charge-discharge cycles obtained from a nicad cell is limited, and in addition, like lead, cadmium compounds are toxic.

Charge capacity of a battery can be defined as useful average current during discharge × time duration of discharge. This follows from the definition of current, $I = Q/t$. Thus, the unit for capacity is the coulomb (C), or ampere-hour (Ah). Car batteries are usually rated in Ah.

Rechargeable lithium cells have a very large *capacity* and are therefore used to power laptop and notebook computers, digital cameras and mobile phones. Recently, rechargeable lithium batteries, on account of their relatively light

weight, large current capability and large electrical energy storage capacity, have been developed to power motor cars.

A variety of secondary and primary cells and batteries are shown in figure 29.25, together with some devices that use them.



Figure 29.25 Some primary and secondary cells and batteries, and some devices that use them.

Practical activity 29.1

Cells and batteries

Identify, in figure 29.25, primary cells, secondary cells, primary batteries, and secondary batteries, giving reasons.

Solar cells (photovoltaic cells)

solar cell | photovoltaic cell

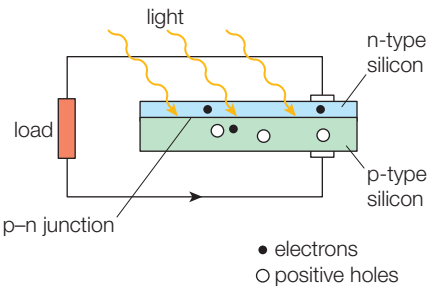


Figure 29.26 A silicon solar cell.

Many calculators are powered by sunlight via **solar** or **photovoltaic cells**. Figure 29.26 shows how a solar cell works. When light shines on the cell, a voltage is generated at the junction of the p-type and n-type materials (see page 475 for a discussion of p-type and n-type semiconductors). This leads to a flow of electrons from the n-type material to the p-type material via the external load. The solar cell converts light energy into electrical energy.

Because solar cells require light to function, they are usually connected to rechargeable storage batteries. Energy converted in daylight is stored in these batteries and can be drawn from them at night.

As well as powering calculators, solar cells are used extensively in spacecraft and satellites for communications equipment.

Comparing cells

Table 29.2 summarises data from some primary cells, secondary cells and solar (photovoltaic) cells.

Table 29.2 Features of different types of cells.

Type of cell	Terminal voltage /V	Nominal internal resistance /Ω	Typical maximum current /A	Rechargeable?	Portability
Leclanché	1.5	~0.5	~5	No	Come in a variety of sizes. Light enough for use in powering TV remote controls, portable radios and flashlights
Alkaline (manganese)	1.6	<0.5	~10	No	Same as for Leclanché cell above
Mercury oxide	1.35			No	Used in the form of small 'buttons', where a steady but small current is required and total weight must be small, e.g. in watches, hearing aids
Nickel-cadmium (nicad)	1.3 fully charged; drops to a steady 1.25 during discharge	<0.1		Yes	Slightly heavier than Leclanché cells of same size. However, can power portable equipment such as motorised toys, and boomboxes, which require larger currents. Some vehicles used in warehouses also make use of large nicad batteries
Lithium	3.5			Yes/no	Light and have large capacity, and so used to power portable computers, and digital cameras. Used as back-up power for computer memories
Lead-acid	2	<0.01	400	Yes	Very heavy and therefore used in stationary applications or in motor vehicles. Have large energy storage capacity and are able to deliver large currents for a long time
Solar	0.7 in bright sunlight			Require light to function	Need large surface area in order to generate large amounts of energy. Very light and can be used in portable equipment, such as calculators, where small voltages and currents are required. In space, where size does not present a problem, used for powering communications equipment

Conserving electrical energy

non-renewable ›

A non-renewable energy resource is one which is not replenished in an average lifetime.

The Caribbean is heavily dependent on fossil fuel to meet its power generation needs. Since fossil fuel is a **non-renewable** energy resource (page 141), power companies and consumers need to take steps to reduce wastage of electrical energy. Further, since power wastage leads to a greater demand for electricity, it increases pollution (page 539), as more fossil fuel is burnt.

Power companies reduce power loss by sending electricity at high voltages (and therefore at low currents), thus reducing Joule heating losses. Consumers can reduce wastage by turning on lights and appliances only when needed. Consumers can also use fluorescent or LED rather than incandescent lamps, as fluorescent and LED lamps use less energy to produce the same brightness. Refrigerators and air-conditioned rooms should be well sealed so that less energy can enter from the environment, and hence less energy is needed for cooling. Energy-efficient appliances (e.g. those labelled 'Energy Star') should also be used as these do the same job with less energy converted.

The use of alternative sources of energy, such as direct sunlight for hot water systems (page 217), solar energy via photovoltaic solar cells (page 451), wind power (pages 142 and 451) and hydroelectricity (page 452), will also help to lessen the dependence on electricity generated from fossil fuels.

Internet search term: alternative energy sources in the Caribbean

Chapter summary

- Fossil fuels are the main source of energy for power generation in the Caribbean. Wind, solar and hydroelectric power generation are used on a much smaller scale.
- Electrical power is most effectively distributed using alternating currents.
- Alternating voltages are stepped up by transformers at the generating plant so that power is transmitted at low currents. Two wires of low resistance, generally large-diameter aluminium wires, are used. The combination of low current and low-resistance reduces power loss along the wires.
- The high voltages are stepped down by transformers before use.
- For safety, one of the two wires taking power to a home is earthed at a substation. This is called the neutral wire. The other is called the live wire.
- At the site where the power is used, a third wire, the earth wire, is connected to earth. At its other end the earth wire should be connected to the metal casing of appliances such as stoves, for further protection of the user against electric shocks.
- Fuses and circuit-breakers in electrical circuits protect against short circuits and overloads.
- In household electrical circuits, loads (e.g. lamps) are wired in parallel. Electrical outlets are usually wired in a ring circuit, which allows twice the power (compared with a simple parallel arrangement) to be delivered safely using the same size of wire. Fuses, circuit-breakers and switches are placed in the live line in series with loads.
- Voltage fluctuations in the form of surges, spikes or low voltages can cause much damage to electronic equipment, and motors. Any electrically operated equipment, as well as the wires connected to them, can become overheated due to the large currents associated with high voltages.
- Power companies use kilowatt-hour (kWh) as the unit of electrical energy.
- Cells allow portability of electrical devices. Some are primary cells (non-rechargeable) and others are secondary cells (rechargeable).
- Cells are used for different purposes depending on their capacity (how much charge they can hold), e.m.f., size of deliverable current, shelf life and steadiness of voltage during discharge.
- Wastage of electrical energy not only results in higher electricity bills but also hastens the depletion of the world's fossil fuel supplies and aggravates pollution to the environment.
- Ways of conserving energy, and therefore fossil fuel supplies, include: turning off equipment not being used; using energy-efficient devices, e.g. fluorescent rather than incandescent lamps; ensuring that appliances such a refrigerators and air-conditioners are properly installed and in proper working order; and making use of alternative sources of energy.

Answers to ITQs

ITQ1 The conveyor belt; the till and price readout; the bar-code reader; the till receipt printer.

ITQ2 Suriname (at Lake Brokopondo).

ITQ3 The energy from the Sun causes the water to evaporate. The water vapour rises and condenses in the clouds to form water droplets. Rain falls from the clouds on to these high regions. Thus, the gravitational potential energy stored by the water in the high regions originally came from the Sun.

ITQ4 $P = IV$

(i) $\frac{I_1}{V_1} = P_1 = \frac{100\,000\,\text{W}}{1000\,\text{V}} = 100\,\text{A}$

Power lost $= I_1^2R = 100^2 \times 2 = 20\,000\,\text{W}$

(ii) $\frac{I_2}{V_2} = P_2 = \frac{100\,000\,\text{W}}{10\,000\,\text{V}} = 10\,\text{A}$

Power lost $= I_2^2R = 10^2 \times 2 = 200\,\text{W}$

ITQ5 A generator can be taken off-line for servicing while the others continue to supply the electricity needs of the system.

ITQ6 **(i)** Both A and B; **(ii)** neither; **(iii)** neither.

ITQ7 **(i)** Both A and B; **(ii)** B only; **(iii)** A only.

ITQ8 The current will have two parallel paths to earth – the path through the consumer's body and that through the earth wire. Since the earth wire has practically zero resistance and the consumer's body resistance is appreciable, the earth wire will take almost the whole current that passes to earth.

ITQ9 A, Circuit breaker; B, main switch.

ITQ10 13 A

ITQ11 A thick jelly-like paste is less likely to spill from a battery than a liquid.

ITQ12 The depolariser, manganese(IV) oxide, is an insulator. The powdered carbon mixed with the manganese(IV) oxide helps to increase the conduction of electrons across it.

ITQ13 The heat of the Sun probably causes the hydrogen gas to rise from the carbon electrode. Tapping might also produce a similar effect. Once the insulating layer of hydrogen gas is removed, electrons flow readily from the carbon electrode and the cell delivers current. Freezing probably causes the hydrogen bubbles to decrease in volume, reducing the insulating layer caused by them.

ITQ14 Six (1.5 V × 6 = 9 V)

Examination-style questions

Items 1–4

The following are four types of power generation:

A fossil fuel **B** hydroelectric **C** wind **D** solar

1 Which type of power generation is *most* polluting?

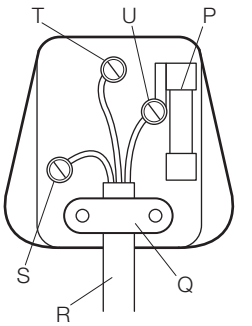
2 Which type of power generation makes use of a chemical reaction?

3 Which type of power generation is *most* used in the Caribbean?

4 Which type of power generation makes use of stored gravitational energy?

Items 5–7

The diagram shows a three-pin 13A fused plug.



5 Which of the labelled items is the fuse?

A P C R

B Q D S

6 Which item is used to anchor the neutral wire?

A Q C T

B S D U

7 What size of fuse should be used in the plug if it is to be connected to a lamp marked 110V, 100W?

A 1A C 5A

B 2A D 13A

8 Generating stations send power over long distances using:

A high voltage, high current
B high voltage, low current
C low voltage, high current
D low voltage, low current

9 An electricity company's fixed rate is 20¢ per unit and the fuel adjustment rate is 0.3¢ per unit. Calculate the cost of operating eight 300 W floodlights for 10 hours.

10 An electric stove has three stove-top heating elements of resistance 48 Ω each and one of resistance 24 Ω. The elements can be switched on and off independently.

(i) Draw a circuit diagram (using the symbol for a resistor as that for a heating element) to show how the elements and switches are connected to the live and neutral wires of a 240 V supply.

(ii) Estimate the total current drawn from the supply when all four elements are operating at full power.

11 The 25 kV (25 000 V) output of a certain power station is first stepped up to 200 kV by a device X.

(i) What is the name given to the device X?

(ii) If the power station delivers 4 MW (4 000 000 W) of power to X, calculate:

(a) the input current to X;

(b) the output current from X (assuming no power loss in X).

(iii) Give a reason why, when transmission over long distances is required, power is sent at high rather than low voltages.

12 What is the main advantage of using ring circuits rather than simple parallel circuits for electrical outlets in homes?

13 (i) State the differences between a primary cell and a secondary cell.

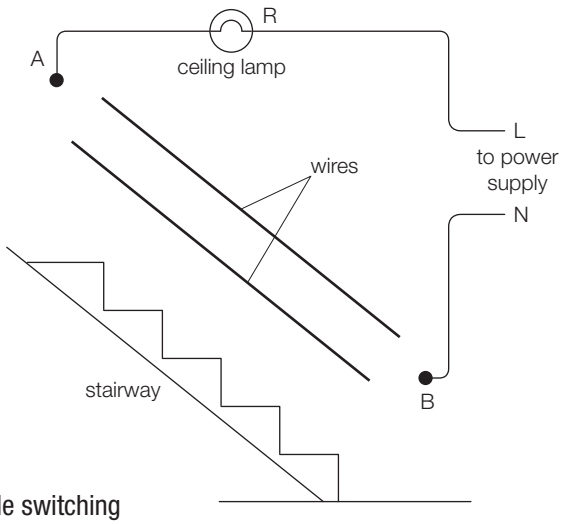
(ii) Discuss the features of a primary cell that affect:

(a) its life;

(b) its e.m.f. (voltage);

(c) the size of the current it can deliver under load.

14 The diagram shows a lamp R, which must be switched on or off from position A as well as position B.



(i) On the diagram, draw a suitable switching arrangement that could accomplish this.

(ii) Name the type of switches used.

15 State *one* advantage and *one* disadvantage in each case when the following type of battery is used in the given application:

(i) Lead-acid batteries in cars;

(ii) nickel-cadmium batteries in motorised toys.

30 Electronic Devices

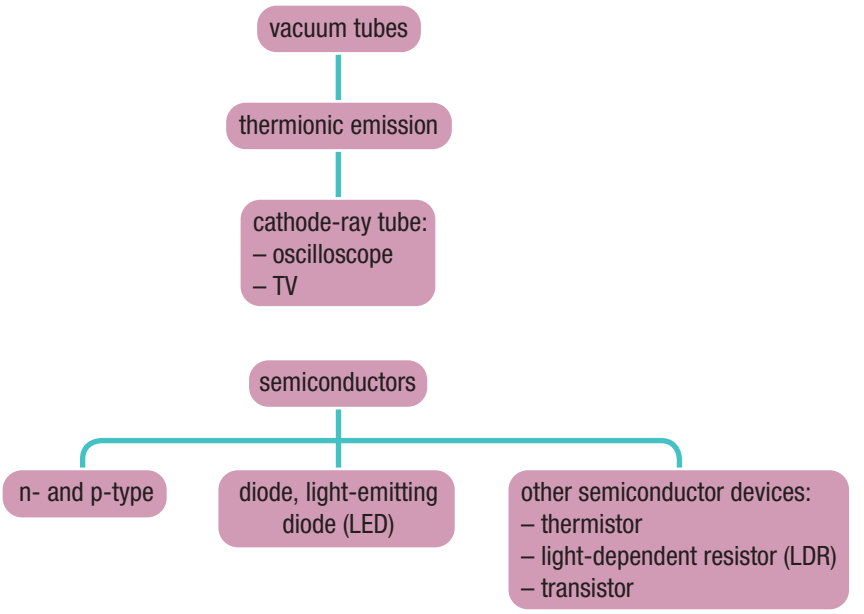
By the end of this chapter you should be able to:

- state what is meant by thermionic emission
- describe the functioning of a cathode-ray tube and uses to which cathode-ray tubes are put
- represent both direct current (d.c.) and alternating current (a.c.) on current–time or voltage–time graphs
- use current–time or voltage–time graphs to find the period and frequency of alternating currents or voltages
- describe a semiconductor in terms of charge carriers and resistance
- draw a current vs. voltage characteristic (graph) to show the behaviour of a p-n junction
- describe how a semiconductor diode can be used in half-wave rectification
- sketch voltage–time graphs to compare the variation of voltage with time before and after rectification
- distinguish between direct current from batteries and rectified alternating current by consideration of their voltage–time graphs
- devise a simple test to determine whether a semiconductor diode is defective
- describe the behaviour of semiconductor devices such as thermistors and light-dependent resistors (LDRs)



Figure 30.1 A computer playing chess. The computer remembers all its opponent's moves – good and bad!

Meet today's computer. The computer of today can 'talk', can convert speech into print as you talk, can fly an aircraft, can run a factory, can play chess (see figure 30.1), can forecast weather – and much more! The functioning of the computer is based on electronics.



Introduction: electronics

electronics

Although **electronics** takes its name from the negatively charged electron, there are also positively charged carriers known as 'holes'. Both types of charge carriers, electrons and holes, feature in electronics. We will be studying two types of electronic devices – vacuum tubes and semiconductor devices – that make use of these charge carriers.

Vacuum-tube devices

cathode ray ›

Vacuum-tube electronic devices were developed following the discovery of 'cathode rays' in the late 1800s. **Cathode rays**, which we now know to be a stream of electrons, were demonstrated when a high voltage such as 10 000 V was applied across a gas at very low pressure (see chapter 32, page 498). Experiments showed that cathode rays:

- travelled in straight lines;
- were deflected by a magnetic field; and
- were deflected by an electric field.

Thermionic emission

thermionic emission ›

In vacuum-tube devices nowadays cathode rays are produced by a simpler process called **thermionic emission**. This occurs when metals are heated and some electrons gain enough energy to escape the surface. Thermionic emission takes place readily when, for example, a mixture of barium oxide and strontium oxide is heated in a vacuum tube, as shown in figure 30.2. When the electrons are liberated from the heated substance, they are easily attracted by positive voltages and form a stream of cathode rays. This is the basis of the

cathode-ray tube ›

cathode-ray tube.

Cathode-ray oscilloscope (CRO)

Figure 30.3 shows the parts of a cathode-ray oscilloscope (CRO) which makes use of the cathode ray tube.

The heating coil obtains its current from a 6 V supply.

The cathode is a metal surface, which is connected to the negative of a high-voltage supply. The surface is coated with an oxide, which, when heated by the heating coil, gives off electrons by thermionic emission.

The three hollow metal cylinders connected to high positive voltages are called anodes. The high

anode ›

anode voltages accelerate the electrons given off by the cathode to very high speeds so they travel straight through to the screen.

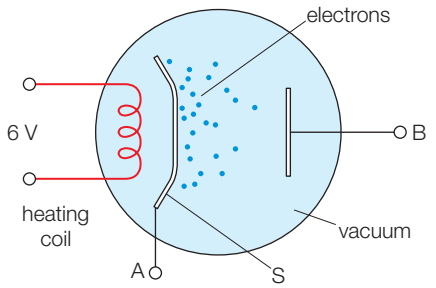
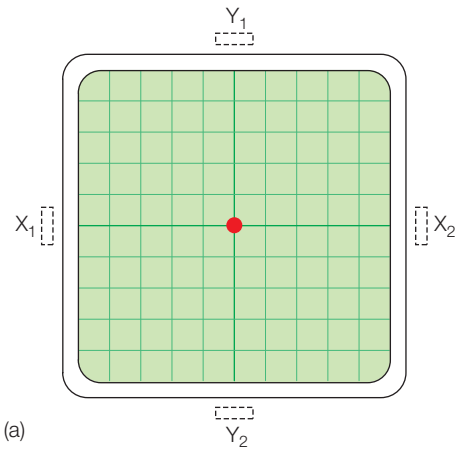


Figure 30.2 A cathode-ray tube. Electrons are leaving the heated surface, S; this is thermionic emission. If B is made positive with respect to A, electrons will travel from A towards B.



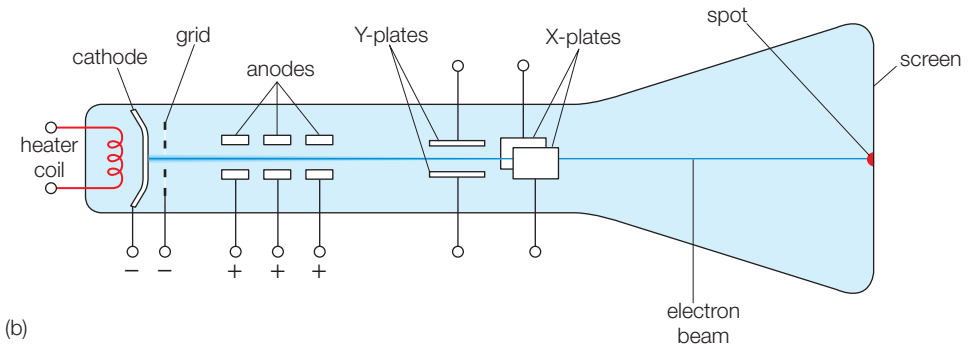


Figure 30.3 (a) Cathode-ray tube of an oscilloscope. (b) Front view of cathode-ray oscilloscope tube. The relative positions of the X- and Y-deflection plates (X_1X_2 and Y_1Y_2) are shown.

The three anodes operate at slightly different voltages. By varying the voltage at the centre anode, the beam of electrons can be focused to a tiny spot on the screen.

phosphor ›

The screen is coated with a **phosphor**, a substance that gives off light when bombarded with electrons.

When a voltage is applied to the pair of plates, Y_1Y_2, the electron beam is deflected. For example, if Y_1 is made negative, the beam deflects in a downward direction, that is along the negative *Y*-axis. The plates are therefore called the Y-**deflection plates** (or Y-plates).

deflection plates ›

X_1 and X_2 are the X-deflection plates (or X-plates), and if X_1 is made negative the beam deflects to the right. Thus, by controlling the voltages across Y_1Y_2 and X_1X_2, the spot can be moved to any position on the screen.

grid ›

The **grid** in front of the cathode is a mesh-type electrode whose voltage can be varied. When the grid is made negative, some of the slower-moving electrons leaving the cathode are repelled back. If the grid is made more negative, even the faster-moving electrons are repelled back towards the cathode. Thus, by controlling the negative voltage applied to the grid, we can vary the strength of the electron beam that reaches the screen, and hence the brightness of the spot.

ITQ1

In what direction would the electron beam be deflected if Y_1 was made positive?

Using the cathode-ray oscilloscope

X-shift and Y-shift controls

When the oscilloscope is turned on, the spot might not appear at the centre of the screen. It can be centred by adjusting the X-shift and Y-shift controls shown in figure 30.4. These vary the voltages applied across the X- and Y-plates, causing the spot to shift in either the X or Y direction to the desired zero position.

Focus control

A spot that is out of focus (looking blurred) can be made sharper by adjusting the focus control. The adjustment causes voltage changes in one of the anodes that control the sharpness of the spot seen on the screen.

Brilliance (brightness or intensity) control

The brilliance or brightness control adjusts the voltage applied to the grid. The brightness can be decreased by applying a grid voltage that is very negative.

Y-input terminals

The voltages being studied are connected to the Y-input terminals of the oscilloscope, which in turn are connected to the Y-plates. Voltages connected across the Y-input cause the spot on the screen to move up or down. If an alternating voltage is applied to the Y-input at a frequency of 30 Hz or more, the spot moves up and down very quickly at the same frequency. At this rapid rate, the spot looks like a vertical line (figure 30.5).

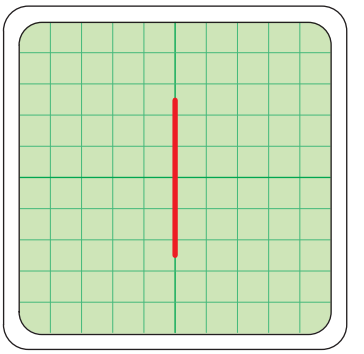


Figure 30.5 When an alternating voltage of frequency of 30 Hz or more is applied to the Y-plates, a vertical line is seen on the screen.

Y-gain control

This control is calibrated, usually in volts per cm, to indicate the size of the voltage applied across the Y-plates. If the gain control is set at 5.0 V per

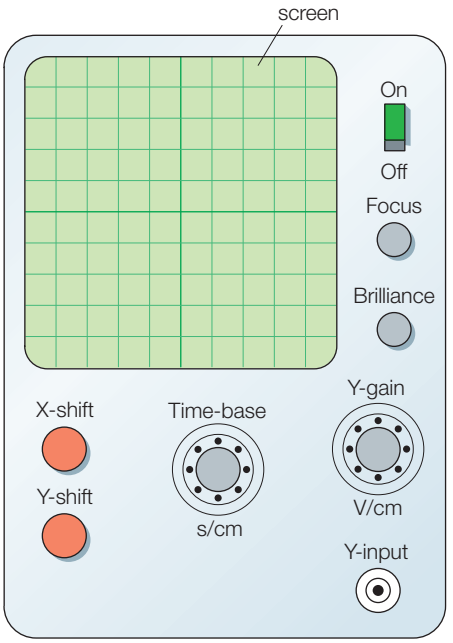


Figure 30.4 The front panel of an oscilloscope.

ITQ2

If the gain control is set at 0.02 V cm^{-1} and the spot moves to the 3.0 cm mark, what is the applied voltage?

centimetre, for example, and an applied d.c. voltage causes the spot to move to the 2.0 cm mark, as shown in figure 30.6, then the applied voltage is 2.0 × 5.0 V, that is 10 V.

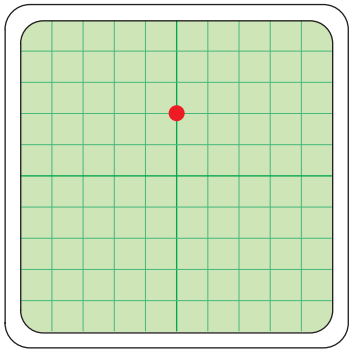


Figure 30.6 Appearance of the spot on the screen when the gain is set at 5 V cm^{-1} and d.c. voltage is 10 V.

Time-base control

time-base

A '**time-base**' voltage is usually applied to the X-deflection plates in order to show how the voltage across the Y-deflection plates is varying with time. The time-base voltage (generated by circuits inside the oscilloscope) increases linearly with time, which causes the spot to move at a steady speed from the X_1 end of the screen towards the X_2 end. This steady movement of the spot across the screen is called a **sweep**.

sweep

The voltage applied to the X-deflection plates is then rapidly decreased, so that the spot quickly returns to its starting point. This swift return movement is called the **fly-back**.

fly-back

The time-base control is usually calibrated in seconds per centimetre (figure 30.7). When the control is set at 0.2 s cm^{-1}, the spot sweeps across 1 cm of the screen in 0.2 seconds. Thus, if the screen is 10 cm wide, the spot takes 10 × 0.2 s or 2 seconds to sweep across the full width of the screen.

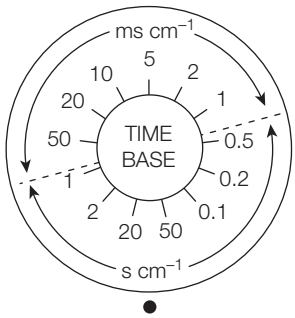


Figure 30.7 Close-up of time-base control.

Worked example 30.1: Time-base voltage set at 2 ms cm^{-1}

How does the spot look on the screen if the time-base voltage is set at 2 ms cm^{-1}? (Assume that there is no voltage across the Y-deflection plates.)

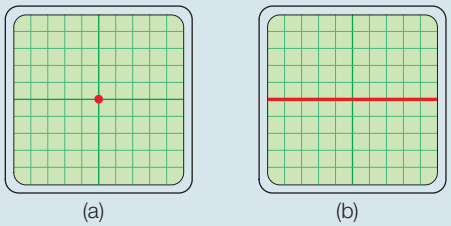


Figure 30.8 Appearance of CRO screen with time-base (a) off and (b) on. (There is no input across the Y-plates.)

Solution

If the time-base voltages is set at 2 ms cm^{-1}, i.e. 0.002 s cm^{-1}, and the screen is 10 cm wide, the spot sweeps and flies back every 0.02 s. At this high speed, the spot would look like a line on the screen during each sweep (figure 30.8 (b)). The sweep frequency would be 1/0.02 sweeps per second or 50 Hz. The line would not appear to flicker since the eye perceives the rapid succession of superimposed lines at frequencies greater than 30 Hz as one steady line.

ITQ3

At what value must the time-base control be set so that the spot sweeps the entire 10 cm screen in 0.05 s?

Some uses of the oscilloscope

Measuring voltage

The height of the spot above or below the centre of the screen gives the voltage being applied to the Y-deflection plates via the input terminals. The d.c. voltage shown in figure 30.6 is 10 V. An alternating voltage is shown in figure 30.5. The **peak voltage** (that is, the voltage from the centre to the maximum) is 12.5 V, assuming a gain control setting of 5 V cm^{-1}. The peak-to-peak voltage, or voltage from minimum to maximum, is 25 V.

peak voltage

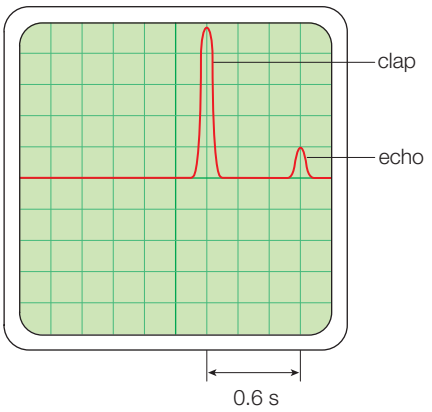


Figure 30.9 Oscilloscope display showing a time interval measurement between two pulses.

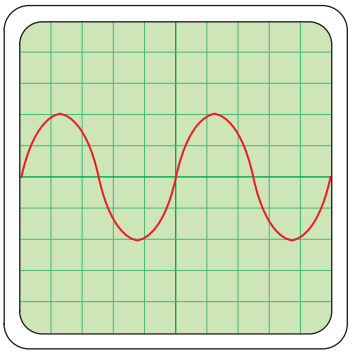


Figure 30.10 The oscilloscope screen shows two cycles of an alternating voltage. The time-base is set at 0.002 s cm^{-1}.

MATHEMATICS: sine curve

IT Q4

If the time-base in figure 30.10 was set at 0.005 s cm^{-1}, determine (i) the period, and (ii) the frequency of the voltage applied to the Y-plates.

The CRO is a very good voltmeter. It has a very high internal resistance and takes almost no current from the voltage source being measured (see page 408).

Measuring time and frequency

The horizontal distance moved by the spot is a measure of time. Thus a CRO can be used, for example, to help measure the speed of sound.

A microphone is connected to the Y-input of an oscilloscope. The time-base is set at 0.2 s cm^{-1}. When a loud clap is made near the microphone, a large voltage pulse generated by the microphone appears on the CRO screen (figure 30.9) as the spot sweeps across the screen. An echo of the clap is reflected from a wall some distance away. The microphone picks up the echo of the clap, and a voltage pulse from the microphone registers on the CRO screen at 3.0 cm from the first pulse. The time interval between the two pulses is therefore 0.60 s. This corresponds to the time taken for the sound to travel to the wall and back.

Worked example 30.2: Speed of sound in air

A wall is 100 m from a second source, and sound takes 0.60 s to travel in air from the source to the wall and back. What is the speed of sound?

Solution

Distance travelled by sound, $d = 2 \times 100\,\text{m} = 200\,\text{m}$

Time taken by sound, $t = 0.60\,\text{s}$

Speed of sound, $v = ?$

$$v = \frac{d}{t}$$
$$= \frac{200\,\text{m}}{0.60\,\text{s}}$$
$$= 330\,\text{m s}^{-1} \text{ (to 2 sig. fig.)}$$

When an alternating voltage (e.g. a sinusoidal varying voltage) is applied to the Y-plates, the spot moves up and down (see page 470). If, at the same time, the spot is made to move horizontally (by the time-base circuit), a sinusoidal wave trace is obtained on the screen.

Figure 30.10 shows an oscilloscope trace produced when an alternating voltage is applied to the Y-input and the time-base is set at 0.002 s cm^{-1}. Two complete cycles of a wavy trace appear on the screen. This indicates that, as the spot sweeps across the screen, the alternating voltage causes the spot to move up and down twice. Since the trace occupies 10 cm, time taken to trace the two cycles = 10 × 0.002 s = 0.02 s

Therefore

time for one cycle = 0.01 s

period ›
frequency ›

In other words, the **period** T of the alternating voltage is 0.01 s. Since **frequency** $f = \frac{1}{T}$,

frequency of the voltage = $\frac{1}{0.01}$ = 100 Hz

Displaying waveforms

The sound from a guitar is different from that of a piano, even if both instruments are playing the same musical note. The sounds can be converted into electrical voltages by a microphone. If the time-base is switched on and the microphone is connected to the Y-input, wave traces that are characteristic

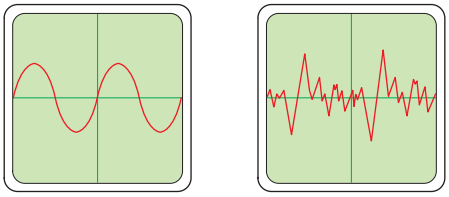


Figure 30.11 Oscilloscope display of wave traces from: (a) a tuning fork sounding a pure note; (b) a violin.

for each instrument are obtained on the screen. For the same note sounded by each instrument, the fundamental frequency is the same; both wave traces show the same number of cycles in a given time. However, the wave shapes are different, indicating that each instrument's sound consists of 'overtones' characteristic or that instrument only (figure 30.11).

Similarly, the voices of different individuals can be displayed on an oscilloscope screen and the differences in the wave shapes produced by each person can be analysed. In hospitals, wave shapes on a CRO screen can give information to the doctor about possible damage to the heart. Apart from displaying waveforms, cathode-ray tubes are also used in air traffic control, to display the position of aircraft being tracked by radar (figure 30.12).



Figure 30.12 Positions of aircraft shown on CRO radar screens.

Displaying TV pictures

In TV picture tubes, the electron beam is deflected, not by X- and Y-plates, but by four 'magnetic deflection' coils carrying current. This coil arrangement can bend the beam much more than the plate arrangement used in the oscilloscope tube, thus producing a larger picture using a shorter tube. TV tubes are now being replaced by 'flat-screen' devices which make use of semiconductors.

Semiconductors

Semiconductor properties

Semiconductors is the name given to substances that appear in the Periodic Table between conductors and insulators. Two examples are germanium and silicon.

Each atom of a semiconductor (figure 30.13) has four outermost electrons (called valence electrons). By covalent sharing of electrons with neighbouring atoms, each atom obtains a stable arrangement of eight outer electrons. At room temperature, however, some electrons gain enough thermal energy to be dislodged from the parent atoms. Vacancies called '**holes**' are left when electrons become free. Since the atom as a whole is neutral when all its electrons are present, the 'holes' behave as though they have a positive charge.

CHEMISTRY: Periodic Table
hole ›

It takes energy to dislodge an electron from its parent atom, resulting in a free electron and a hole. The energy may be thermal, due to vibration of the atoms of the semiconductor. The energy may also come from electromagnetic radiation (such as visible light or infrared radiation) or from moving particles (such as beta particles).

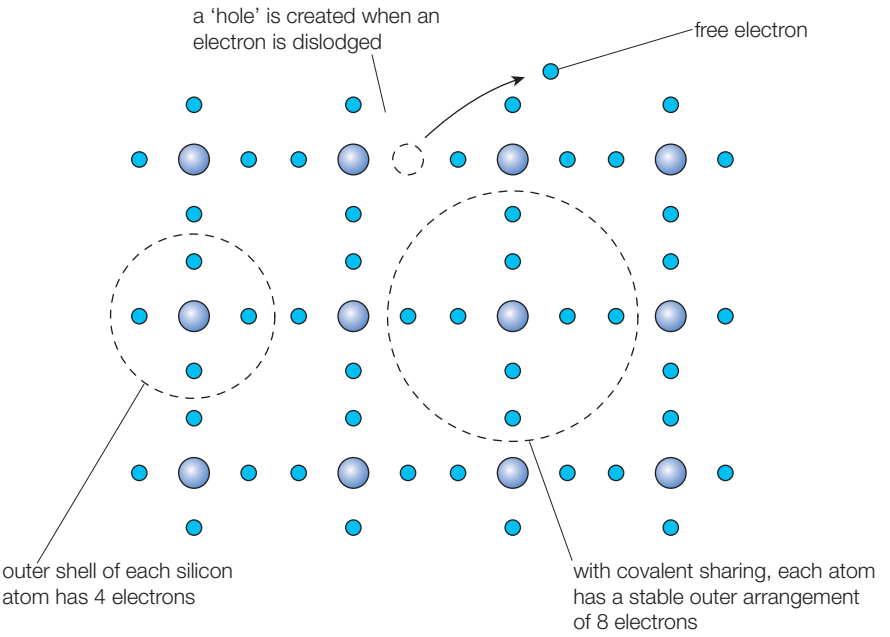


Figure 30.13 Atomic structure of a semiconductor element showing outermost electrons.

At room temperature, semiconductors are relatively poor conductors. This is because, compared with conductors, semiconductors do not have many 'free' electrons for conduction. As temperature rises, however, more electrons become dislodged from atoms and become available for conduction. Hence, the resistance of semiconductors decreases with increasing temperature. This is quite unlike the case with conductors, where resistance increases as temperature goes up.

Electrons and holes

Each time an electron is dislodged from a filled shell, a hole appears. In a pure semiconductor, there are therefore equal numbers of electrons and holes. If a battery is connected to the ends of the semiconductor (figure 30.14), 'free' electrons move away from the end connected to the negative terminal. A valence electron from atom A may gain sufficient energy to be dislodged. This electron may move towards B and fill the hole at B. This leaves a hole at A. Since the hole that was at B is now at A, it is as if the hole has moved from B to A. Thus the current in a semiconductor seems to consist of free negatively charged electrons moving towards the positive terminal and positively charged holes moving (in the conventional current direction) towards the negative terminal; two types of charge carriers are involved.

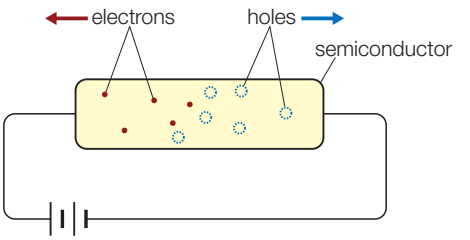


Figure 30.14 Battery connected to a semiconductor. Note the direction of flow of the electrons and the holes.

Effect of impurities (doping)

intrinsic semiconductor ›

Semiconductors in which holes and free electrons are produced as a result of atoms gaining energy are called '**intrinsic semiconductors**'. However, the conducting properties of a semiconductor can be altered considerably by

doping ›

introducing very small traces of impurities during manufacture, a process known as '**doping**'.

Typically, each atom of a semiconductor has four valence electrons. Supposing an atom of phosphorus, which has five valence electrons, is introduced into the structure, as shown in figure 30.15. There is now a 'free' electron, and the semiconductor now has more electrons than holes. When a semiconducting material is doped with a pentavalent element, the great majority of the charge carriers will be (negatively charged) electrons, and the material is called an '**n-type**' semiconductor.

The impurities added to semiconductors are called 'dopants'.

Pentavalent means having five valence electrons.

Trivalent means having three valence electrons.

n-type ›

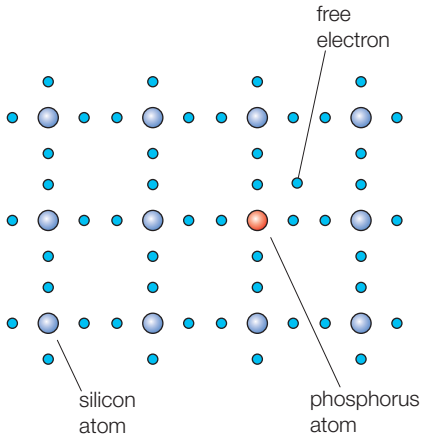


Figure 30.15 Formation of an n-type semiconductor.

During the manufacturing process, the semiconducting material may be doped with an element that has three valence electrons (trivalent). Large numbers of positively charged holes are then formed in the structure (figure 30.16). Semiconductors with more holes than electrons are called '**p-type**' semiconductors.

p-type ›

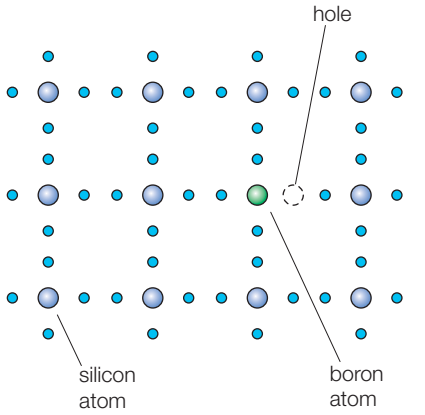


Figure 30.16 Formation of a p-type semiconductor.

The overall effect of adding impurities is to increase the number of charge carriers, resulting in increased conduction. The addition of even small traces of impurity increases conduction dramatically, and so, during the manufacture of semiconductors, great care is taken to exclude impurities or to precisely control the amount entering the semiconductor.

Semiconductors in which conduction occurs because of the presence of doped impurities are called '**extrinsic semiconductors**'.

extrinsic semiconductor ›

The p-n junction diode

A semiconductor can doped to become p-type material on one side and n-type on the other. The meeting of the p-type and n-type materials is called the **p-n junction** (figure 30.17). Since there are two electrodes attached to the semiconductor, one to the p-type material and one to the n-type material, the device is called a diode. The electrode connected to the p-type material is called the anode and that connected to the n-type material, the cathode. The semiconductor is usually encased in plastic for protection and the negative terminal is usually indicated (see figure 30.20).

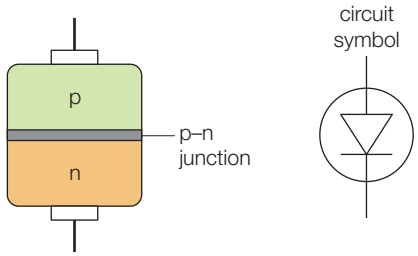


Figure 30.17 A p-n junction and symbol.

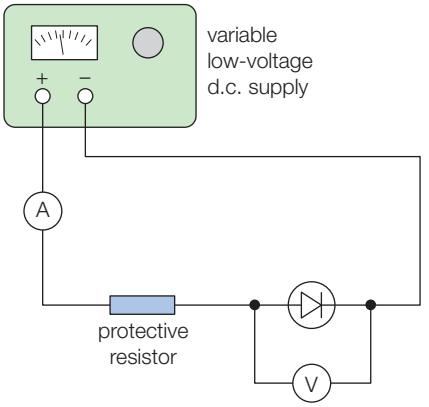


Figure 30.18 Circuit for determining the current–voltage characteristic of a p-n junction diode.

The p-n junction characteristic

The graph showing the relationship between the current, I, produced when a voltage, V, is applied across a p-n junction is called a p-n junction characteristic or a semiconductor diode characteristic. Figure 30.18 shows a circuit that can be used to investigate the current–voltage characteristic of a p-n junction diode. Figure 30.19 shows a typical characteristic.

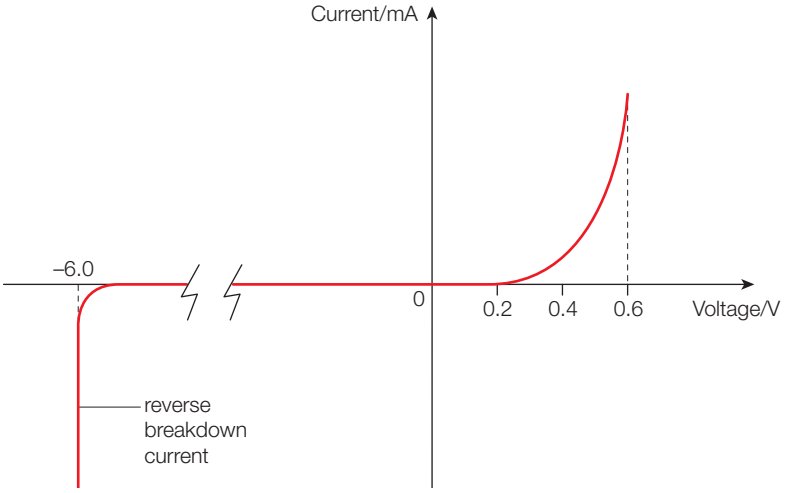


Figure 30.19 Current–voltage characteristic for a p-n junction diode.

A current results when the anode of the diode is connected to the positive of the applied voltage, and the negative end of the diode to the negative terminal of the cell. This is called **forward biasing** of the diode. When the applied voltage connections to the diode are reversed, there is hardly any current in the diode. This is called **reverse biasing**. Thus, in reverse bias, the diode shows a very high resistance; but in forward bias, the resistance shown by the diode is very much smaller.

forward biasing ›

reverse biasing ›

Manufacturers usually indicate the polarities of diode terminals. A band at one end, for example, means that the n-type terminal is at that end (figure 30.20). In other diodes, the longer terminal is connected to the anode, i.e. to the p-type material, and the shorter to the cathode, the n-type material.

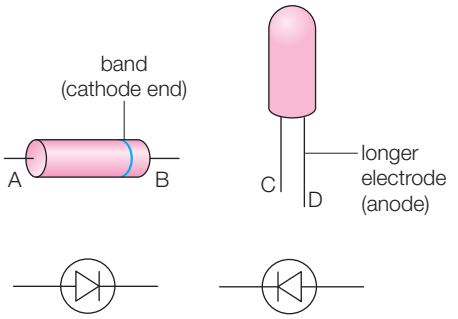


Figure 30.20 Polarity indicators on junction diodes.

As the forward voltage is increased from zero, the forward current increases, at first very slowly and then more quickly and very nearly linearly for a while. The current in the diode causes it to get hot, and the heating up leads to the release of more charge carriers. There is a further increase in current and further heating. Beyond a certain applied forward voltage, the current is so big that the diode burns up. Junction diodes, therefore, cannot withstand high forward voltages.

Since the current in the diode increases faster than the voltage across it, the resistance of the diode decreases with large currents. This means that the diode becomes a better conductor as the voltage across it increases. Resistors are therefore usually placed in series with diodes to limit the current flowing through them and prevent them from destruction by overheating.

Although there is hardly any current (only a few microamperes) through the diode when reverse-biased, if a certain reverse voltage is reached, there is a large and sudden 'breakdown' current, which could also destroy the diode by overheating.

Not all junction diodes are destroyed by large currents. Specially constructed power diodes are able to withstand currents of several amperes.

ITQ5

To which labelled terminal in each of the diodes shown in figure 30.20 must the positive of a voltage supply be connected in the forward-bias mode?

Practical activity 30.1

Testing a diode

Caution: Make sure you show your plan (including your circuit diagram) to your teacher before attempting to carry out the investigation practically.



You are to design and carry out a simple test to find out which of three p-n junction diodes is/are defective (one may be 'open', one 'shorted' and the third 'good'). (Hint: You must first define the behaviours 'good' and 'defective' based on your understanding of diode characteristics.)

ITQ6

Why must a low-voltage supply be used when testing a semiconductor diode?

Rectification using a junction diode

Figure 30.21 shows a 1 kΩ resistor connected in series with a diode and an a.c. supply.

When the Y-input of the oscilloscope is connected across a suitable resistor that is in series with an a.c. source only, a trace like that in figure 30.22 (a) appears on the screen. The trace shows that the supply voltage is positive for half a cycle and negative for the next half-cycle. The current therefore *reverses direction* halfway through every cycle.

When the Y-input of the oscilloscope is connected across the 1 kΩ resistor, a trace like that in figure 30.22 (b) appears. The trace shows that the voltage across the resistor is positive for half a cycle and zero for the next half-cycle. Thus, the current is flowing only one way through the resistor. The current is therefore direct current (d.c.), because it flows in one direction only. However, it flows in on-off pulses.

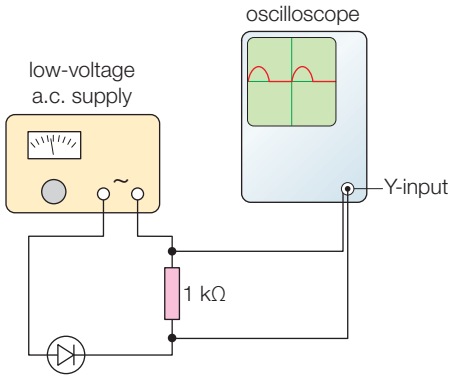


Figure 30.21 Diode connected in series with a 1 kW resistor and an a.c. supply.

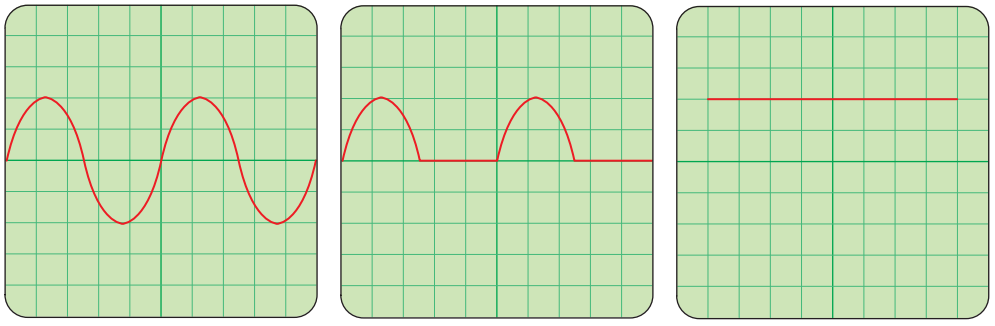


Figure 30.22 Traces obtained on oscilloscope: (a) a.c. voltage; (b) pulsating direct current (d.c.); (c) steady d.c.

For comparison, the d.c. voltage from a battery connected to the Y-input of the oscilloscope is shown in figure 30.22 (c). The d.c. voltage from the battery is steady, whereas the d.c. voltage across the resistor in the diode circuit is **pulsating**.

pulsating

A junction diode can therefore convert alternating current to direct current. This is called **rectification**. Diodes made for this purpose are sometimes called **rectifiers.**

rectification

rectifier

Light-emitting diode (LED)

light-emitting diode (LED)

Some specially constructed junction diodes, called **light-emitting diodes** or LEDs, give off light when a current flows through them (figure 30.23). Thus an LED can be used to show the presence of current, and therefore voltage. LEDs are used as power indicators in electrical equipment. A typical LED operates safely at a little less than 2 V. The current flowing is then about 10 mA. An example of how to protect an LED is given in chapter 25 (page 393).

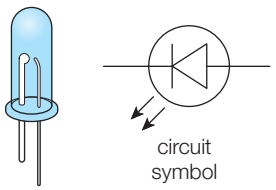


Figure 30.23 An LED and its symbol.

ITQ7

What protective resistance is needed for the safe operation of an LED from a 6 V battery?

Semiconductor 'resistors'

Thermistor

When a semiconductor is heated, the increase in thermal energy causes more valence electrons to be dislodged from their parent atoms. Since more charge carriers are available in the heated semiconductor, it conducts electricity more easily. Thus, the resistance of the semiconductor decreases with rise in temperature. **Thermistors**, whose resistances are very sensitive to temperature changes, can be used to make simple electronic thermometers (see figure 30.24). As resistance changes, a change in current is seen on the milliammeter. Thermistors can also be used in simple fire-alarm circuits, by replacing the milliammeter in figure 30.24 with an inexpensive piezoelectric buzzer. The buzzer makes a loud sound when the current through it is large.

thermistor ›

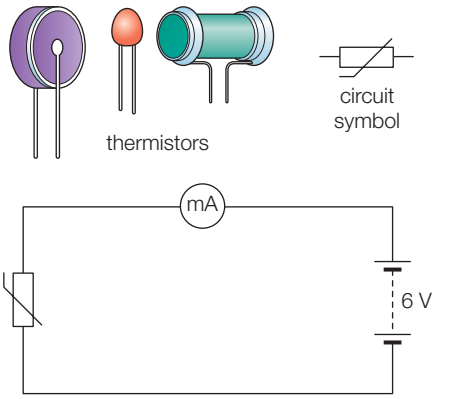


Figure 30.24 Thermistor used in a simple electronic thermometer circuit. The reading on the meter gives an indication of the temperature sensed by the thermistor. Note the circuit symbol for a thermistor.

ITQ8

Would the milliammeter in figure 30.24 indicate a high or a low reading when the thermistor touches a hot object? Explain your answer.

Practical activity 30.2

Thermistor

Using a circuit such as figure 30.24, investigate how the current in a thermistor varies with temperature from about 80°C to about 5°C (you may, for instance, place the thermistor in a beaker of hot water that is allowed to cool, e.g. by the addition of small pieces of ice, and read temperatures as cooling takes place). Record your results suitably and plot a graph. Use your graph to find the average temperature of the members of your group.

Caution: Make sure you show your plan (including your circuit diagram) to your teacher before attempting to carry out the investigation practically.





Light-dependent resistor (LDR)

The resistance of a semiconductor also decreases when light falls upon it. The light energy interacts with the atoms of the semiconductor and dislodges valence electrons; there are more charge carriers and so resistance is lower.

Figure 30.25 shows how a semiconductor **light-dependent resistor (LDR)** can be used in a light-meter circuit. (A light-meter is an instrument for measuring brightness. Photographers use light-meters for measuring brightness of a scene before adjusting their cameras to take pictures.) The LDR is connected in series with a battery and a milliammeter.

light-dependent resistor (LDR) ›

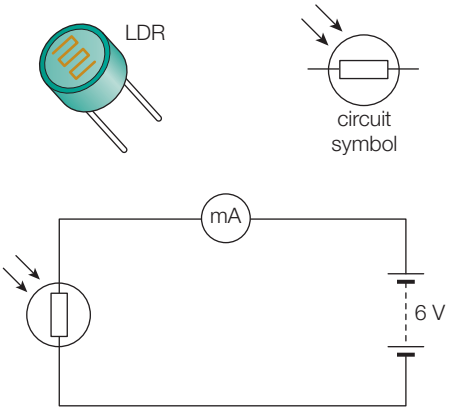


Figure 30.25 A simple light-meter that makes use of a light-dependent resistor (LDR). Note the electrical symbol for an LDR.

ITQ9

Would the milliammeter in figure 30.25 indicate a high or a low reading when bright light illuminates the LDR? Explain your answer.

The resistance of an LDR in the dark can be 10 MΩ (10×10^6 Ω), falling to about 100 Ω in bright light.

Since the resistance of the LDR varies with the intensity of light falling on it, the current in the meter will vary as well. The variation in current in the meter indicates the brightness of a scene. The varying current could also be used to automatically alter the aperture in a camera, depending on the amount of light in the scene being photographed.

Transistor

transistor ›

Transistors are made of three layers of semiconducting material, one type of layer sandwiched between the other two types, e.g. (p-type, n-type, p-type) or (n-type, p-type, n-type). An integrated circuit (IC) may consist of thousands of transistors, diodes and resistors all etched on to a single semiconductor crystal smaller than your fingernail. Integrated circuits are used extensively in computers, calculators and electronic watches. The topics of transistors and integrated circuits fall outside the scope of the syllabus.

Chapter summary

- The emission of electrons from the surface of a substance when it is heated is called thermionic emission. It occurs readily when the substance is heated in a vacuum.
- In a modern cathode-ray tube, electrons given off by thermionic emission are accelerated in a narrow beam by a positive electrode called an anode.
- The cathode-ray tube is used in the oscilloscope and in TV picture tubes.
- Semiconductors are substances whose resistances fall between those of conductors and insulators. The charge carriers in semiconductors are (negative) electrons and (positive) holes.
- In a pure semiconductor there are as many electrons as holes, since a hole is generated whenever a valence electron is dislodged from its parent atom.
- Semiconductor material can be 'doped' with atoms of some pentavalent or trivalent elements. The resulting semiconductor may then have an excess of electrons, and is known as n-type material, or an excess of positive holes, in which case it is p-type.
- A p-n junction is formed by manufacturing p-type and n-type material on the same semiconductor crystal.
- A p-n junction diode is forward-biased when the positive of the voltage supply is connected to the p-type terminal of the diode. It is reverse-biased when the positive of the voltage supply is connected to the n-type terminal of the diode.
- Appreciable current only flows through a p-n type junction diode when it is forward-biased.
- A p-n junction diode can function as a rectifier, converting a.c. into d.c., or as a 'valve'.
- The resistance of semiconductors decreases with application of energy to the semiconductor. This property is used in thermistors and light-dependent resistors (LDRs).

Answers to ITQs

ITQ1 Upward. (The positive plate would attract the negatively charged electrons.)

ITQ2 3.0 cm × 0.02 V cm^{-1} = 0.06 V

ITQ3 The screen is 10 cm wide, therefore,
the spot travels 10 cm in 0.05 s
the spot travels 1 cm in $\frac{0.05}{10\,s}$, i.e. 0.005 s
The time-base control is set at 0.005 s cm^{-1}, i.e. 5 ms cm^{-1}

ITQ4 Time for 1 sweep = 10 cm × 0.005 s cm^{-1}, i.e. 0.05 s.
Since 2 cycles are displayed,
time for 1 cycle = $\frac{0.05\,\text{s}}{2}$ = 0.025 s
The period is 0.025 s
Frequency = $\frac{1}{\text{period}} = \frac{1}{0.025\,\text{s}}$ = 40 Hz

ITQ5 A and D

ITQ6 Too high a voltage would result in a large current that could destroy the diode through overheating.

ITQ7 $V = IR$. For a maximum current of 10 mA.
$6 = 10 \times 10^{-3} \times R$
$R = 600\,\Omega$

ITQ8 A high reading. The resistance of the heated thermistor drops because more charge carriers are being generated. Hence the current in the circuit increases.

ITQ9 High reading. The resistance of the LDR drops because there are more charge carriers. Hence the current in the circuit increases.

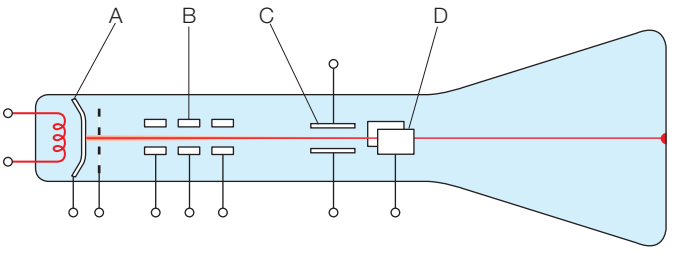
Examination-style questions

1 Thermionic emission is the giving off, from a heated surface, of
- A electrons
- B light
- C protons
- D X-rays

2 The particles that make up the beam in a cathode-ray tube are
- A electrons
- B neutrons
- C photons
- D protons

Items 3 and 4

The diagram shows a cathode-ray tube.

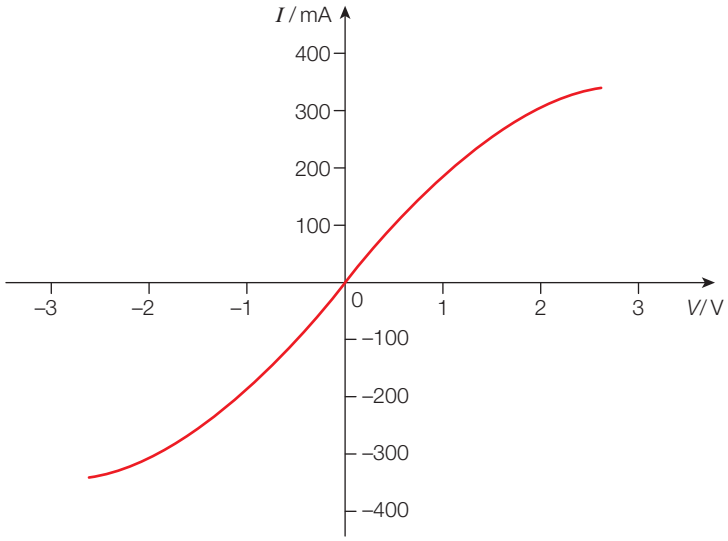


3 Which of the labelled parts represents the Y-plates?

4 Which of the labelled parts represents the cathode?

Items 5–7

The graph shows the variation of current, *I*, with applied voltage, *V*, for a certain device, X.



5 The device, X, is *most* likely to be a:

A diode
B LDR (light-dependent resistor)
C thermistor
D filament lamp

6 At 2.0 V the current is:

A 100 mA
B 200 mA
C 300 mA
D 400 mA

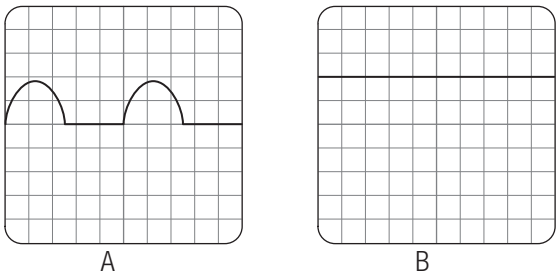
7 The resistance of X is *least* at:

A –1.5 V
B 0.0 V
C 1.0 V
D 2.0 V

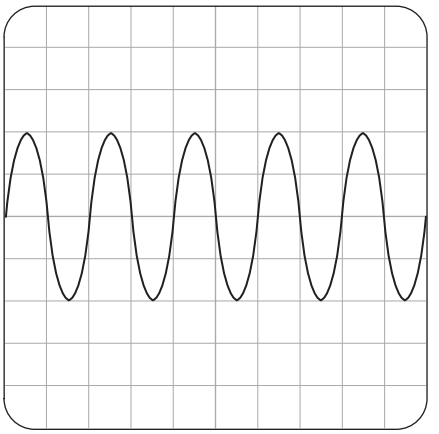
8 The conversion of alternating current (a.c.) to direct current (d.c.) is called:

A amplification
B charging
C doping
D rectification

9 Which oscilloscope trace in the diagram shows *most* closely the voltage–time characteristic of a battery?



10 A loudspeaker emits a pure note, which is picked up by a microphone connected to an oscilloscope. The diagram shows the display obtained on the oscilloscope.



(i) To which terminals of the oscilloscope is the microphone *most* likely to be connected?

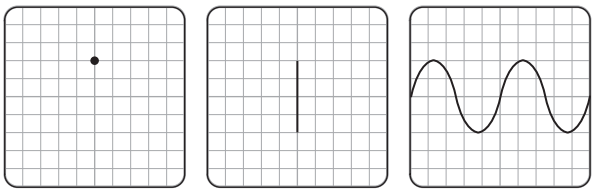
(ii) (a) How many centimetres do *four* complete cycles of the note occupy on the screen?

(b) If the time-base is set at 5 ms cm^{-1}, determine the time taken for *one* complete cycle.

(iii) Determine the frequency of the note coming from the loudspeaker.

(iv) If the speed of sound in air is 330 m s^{-1}, determine the wavelength (in air) of the note coming from the loudspeaker.

11



A B C

Diagram A shows a spot on an oscilloscope screen when the Y-input voltage is 0 volts. The sensitivity is adjusted so that an input of 2 V gives a deflection of 1.0 division on the screen.

(i) Which control should be adjusted to:

(a) bring the spot to the centre of the screen?

(b) make the spot smaller and hence sharper?

(ii) After the adjustment in (i) has been carried out, a voltage, *V*, is applied at the Y-input to give the trace shown in diagram B.

(a) What kind of voltage is *V*?

(b) What is the peak-to-peak voltage of *V*?

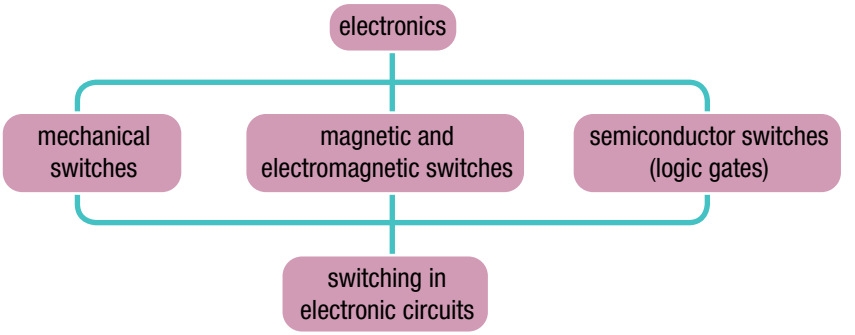
(iii) Which control, if turned on while *V* is applied, will most likely result in the trace shown in diagram C?

(iv) How would the trace in C look on the screen if the sensitivity was adjusted to 1 V for a deflection of 1.0 division?

31 Putting Electronics to Work

By the end of this chapter you should be able to:

- describe the functioning of mechanical and electronic switches
- recall the symbols for AND, OR, NOT, NAND and NOR logic gates
- state the function of each gate with the aid of truth tables
- draw and analyse simple electronic circuits (e.g. alarm circuits and motor control circuits)
- draw and analyse circuit applications involving combinations of not more than three logic gates
- discuss the impact of electronic and technological advances on society



Electronic applications

Putting electronics to work involves using not only electronic devices (such as diodes, thermistors, light-dependent resistors (LDRs) and transistors that were discussed in chapter 30), but also employing appropriate forms of switching.

In this chapter, we discuss mechanical (hand-operated, magnetic and electromagnetic) switches and some simple circuits in which these are used. Next, we examine some electronic 'logic gates' – switches that make use of semiconductors and can act much faster than mechanical switches. As we study these 'logic' circuits, we will also discuss illustrative practical applications, some of which involve a combination of types of switching.

Today, electronic switching circuits have diverse uses such as monitoring and controlling water temperature, turning on a 'night-light' and even providing security for a vault. They have a multitude of other uses, including the production of 'electronic' music on keyboards.



Figure 31.1 A pilot in training. The view through the windows is not real. It is generated by computer. As the pilot adjusts the controls in his or her cockpit, a computer varies the images correspondingly. The computer even bumps and tilts the cockpit to simulate flight in a real aircraft.

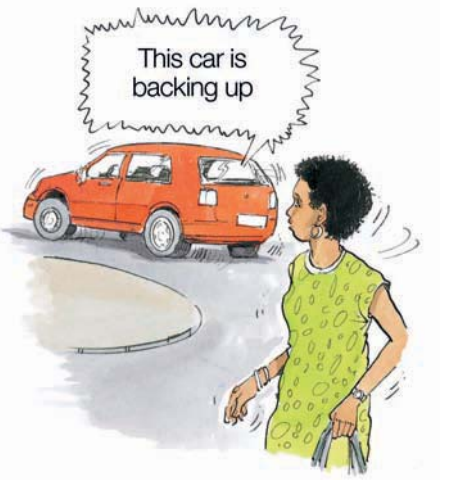


Figure 31.2 Artificial (synthesised) voice. The 'voice' is produced using electronic switches operating at high speeds within electrical circuits, much as digital electronic music of many forms is produced on keyboards.

Computers make use of electronic switches that can go on and off billions of times every second in electronic circuits. The applications of computers seem endless: they are used in almost every field of human endeavour: astronomy, medicine, weather forecasting, business, manufacturing, education and even flight-training and 'voice-synthesis' (figures 31.1 and 31.2).

Switches

Mechanical switches

Mechanical switches control current by opening or closing a circuit. Figure 31.3 shows simple mechanical switches that are operated by hand. Mechanical switches are usually used in series with an electrical supply to turn power on or off.

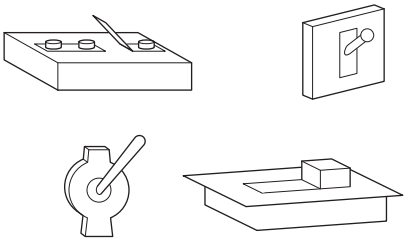


Figure 31.3 Some simple mechanical switches.

Magnetic switches

Mechanical switches can be opened or closed by magnets.

Figure 31.4 (a) shows a 'normally open' (abbreviated 'NO') **magnetic reed switch**. When a magnet is placed beside the switch, magnetism is induced in the thin reeds and they attract each other, closing the contacts. Figure 31.5 shows how the NO switch can be used in a sliding-door burglar alarm circuit.

magnetic reed switch ›

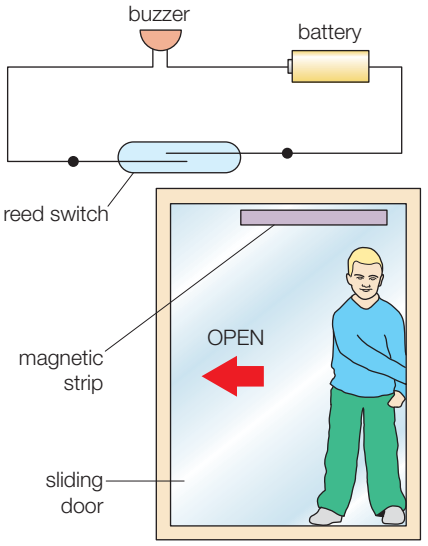


Figure 31.5 A NO reed switch used in a sliding-door burglar alarm.

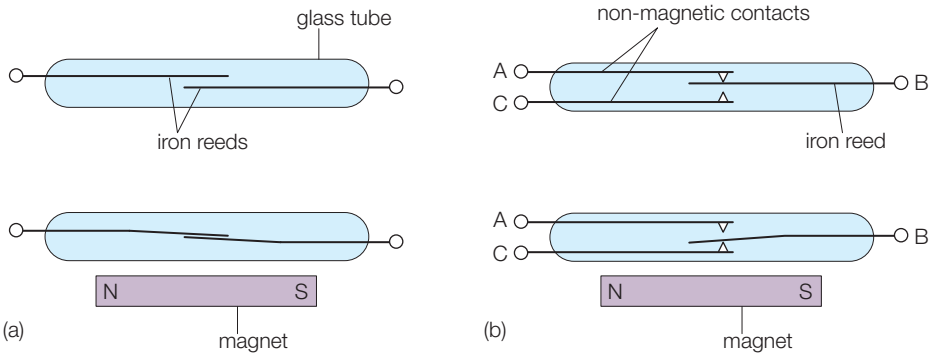


Figure 31.4 (a) Normally open (NO) and (b) normally closed (NC) switches being operated by magnets.

Figure 31.4 (b) shows a 'normally closed' (NC) magnetic reed switch. The path AB is the normally closed path, that is, the position of the switch without a magnet. When a magnet is placed beside the switch, the iron reed is attracted towards the magnet, breaking the AB path but completing the BC path.

Practical activity 31.1

A NC switching circuit

Caution: Make sure you show your plan (including your circuit diagram) to your teacher before attempting to carry out the investigation practically.



Using a magnetic NC switch (figure 31.4 (b)), you are to:

(a) design a circuit that will identify the NC contacts;

(b) design a circuit that will sound an alarm only when a door is opened, and test your design experimentally using a bar magnet placed on a cardboard 'door', or with your teacher's permission, placed on a real door (hint: consider figure 31.5, except that you are now using a NC switch);

(c) predict, and then confirm experimentally, if:

- the orientation of the poles of the magnet matters;
- the circuit will function if the magnet is placed on either side of the switch.

ITQ1

What is a disadvantage of using NO or NC switches as door entry alarms (e.g. in figure 31.5)?

Relay (electromagnetic) switches

Any switch that operates using electromagnetic induction may be called a relay.

reed relay ›
electromagnetic relay ›

A small current in a coil of wire around a reed switch causes the iron reeds to become magnetised and come together (figure 31.6). The small current in the coil switches on a larger current in the reed circuit. This is the principle of operation of the **reed relay**.

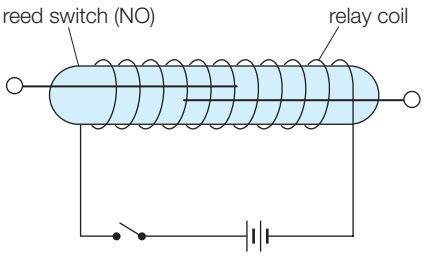


Figure 31.6 Reed relay switch.

The **electromagnetic relay**, shown in figure 31.7, can switch on larger currents than a reed relay. The relay coil of many turns is wrapped around a soft-iron core. When a small current passes through the coil, a powerful electromagnet is produced, which attracts the soft-iron armature (see chapter 27, page 424). The other end of the armature pushes the springy metal and the contacts, C, are brought together. The terminals of the contacts are connected to the circuit to be switched. If the current in the coil becomes zero, the electromagnetic pull on the armature drops to zero. The spring pushes back the armature and the contacts open.

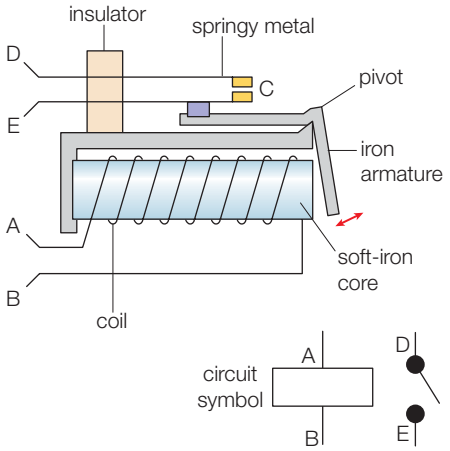


Figure 31.7 Electromagnetic relay switch.

Very high-current circuits, such as starter motor circuits and headlights in cars, are switched by specially constructed electromagnetic relays (figure 31.8) that can handle currents in excess of 60 A.

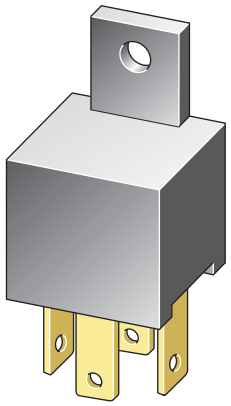


Figure 31.8 A 30 A automotive relay for switching on headlights.

ITQ2

Figure 31.8 shows a relay used for switching on the headlights of cars. Why are there four terminals on the relay?

Electronic switches

A specially constructed type of transistor, called a field-effect-transistor (FET) forms the basis of semiconductor switches commonly used in today's electronic circuits. These switches are called 'gates' because they can open and close a path of current. Electronic switches can act much faster than mechanical switches. High-frequency voltages applied to them can switch them on and off millions of times in one second! This is why electronic switches are used in computer circuits. Also, as electronic switches have no moving parts, they do not wear away as quickly as mechanical switches. On the other hand, mechanical switches can handle much larger currents than electronic switches. We shall discuss semiconductor 'gates' shortly.

Some simple applications of electronics

In all applications described in the rest of this chapter, it is assumed that there is a mechanical switch in series with the power source. In the diagrams, this switch has been omitted for simplicity. As you go through the circuit applications that follow, try to identify mechanical, electromagnetic and/or semiconductor switches.

Circuits based on diode properties

Voltage reversal protection

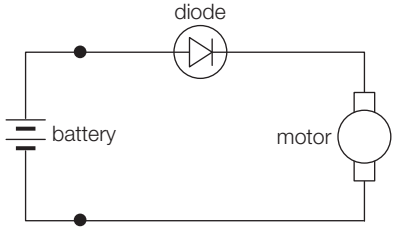


Figure 31.9 Diode used in voltage reversal protection.

Figure 31.9 shows a motor that is required to spin in only one direction when connected to a battery. A power diode is placed in series with the motor. When the battery is connected as shown, current flows and the motor spins. If the battery connections were reversed, the diode would be reverse-biased,

no current would flow and the motor would not spin. This kind of circuit involving a diode in series with a device can also be used to protect electronic equipment that could be damaged by reverse voltages.

Standby power

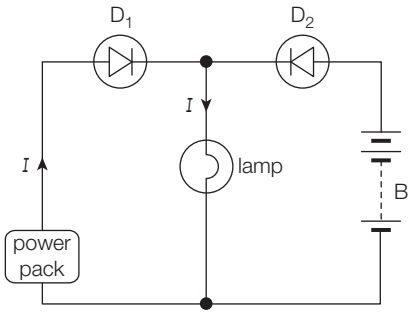


Figure 31.10 Standby lighting circuit making use of diodes.

Exit lamps at public places must always remain lighted. You can make a circuit to demonstrate 'standby' lighting.

A lamp is served by a power pack connected to the mains, as shown in figure 31.10. The voltage supplied by the power pack is slightly greater than that supplied by the standby battery B. Thus a current, *I*, from the power pack circuit lights the lamp. The current, *I*, is blocked by diode D_2 from entering the battery circuit.

Should the mains power fail, the power pack voltage will drop to zero. There will now be current from the battery, B, through diode D_2 and the lamp will light. The current will be blocked by diode D_1 from entering the power pack circuit.

Practical activity 31.2

Two-way switching involving LEDs (light-emitting-diodes)

Caution: Make sure you show your plan (including your circuit diagram) to your teacher before attempting to carry out the investigation practically



You are given a red LED, a similar, but green LED, a 6V battery, two 300Ω protection resistors, and a two-way (single-pole double-throw (SPDT)) switch (i.e. a switch that can make contact with either of two circuits – see figure 29.19).

You are to design and test a 'lavatory door' circuit such that the red 'occupied' light goes on when the switch is in one position and the green 'unoccupied' light goes on when the switch is the other way.

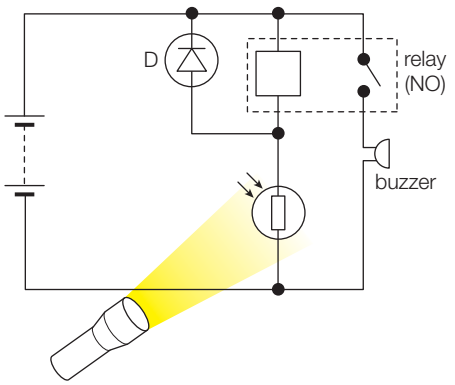


Figure 31.11 Light-operated burglar alarm. When the relay coil is switching off, it develops a large reverse voltage, called a back e.m.f., which can result in a large current in the semiconductor LDR and ruin it. The diode D protects the LDR from damage by allowing this current to bypass the LDR and join the main circuit.

Circuits based on semiconductor 'resistors'

Light-operated burglar alarm

In the dark, the resistance of a light-dependent resistor (LDR) is very high. Thus, very little current flows through the relay coil and LDR from the battery connected as shown in figure 31.11. The relay switch remains open.

When light falls on the LDR, the LDR resistance decreases. A larger current now flows through the coil and LDR from the battery. The relay switch closes and the buzzer sounds the alarm.

This type of alarm can be triggered by a burglar's flashlight in a dark room. The main disadvantage, however, is that the alarm will not stay on once the flashlight is turned off.

Fire alarm

By replacing the LDR in figure 31.11 with a thermistor, the circuit will function as a fire alarm (figure 31.12). When the thermistor gets hot, a large current flows through it because its resistance decreases. The large current now passing through the relay coil turns the relay switch on and the buzzer sounds.

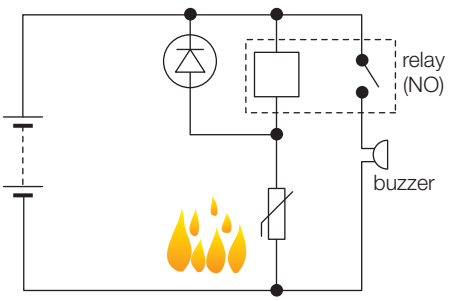


Figure 31.12 Fire alarm circuit involving the use of a thermistor.

Circuits based on electronic logic (gates)

digital ›

Electronic logic circuits make extensive use of electronic switches. The switches in these circuits are operated by voltages in **digital** form. This means that the values of the voltages do not vary gradually; they change only in steps.

binary system ›

The simplest digital system, called a **binary system**, has two states (values) only. These are called logic 1 and logic 0. As shown in figure 31.13, logic 1 corresponds to a closed switch or a 'high' voltage (often 5 V). Logic 0 corresponds to an open switch or a low voltage (often 0 V).

Logic	Logic switch	Voltage (output or input)
1	closed	high
0	open	low

Figure 31.13 Binary logic states.

Logic gates and truth tables

gate ›

truth table ›

Many seemingly complex digital electronic circuits are simply a combination of electronic switches. Each type of combination, called a '**gate**' (figure 31.14), has a characteristic output (or logic state) for given inputs. We use a diagram called a **truth table** to describe how a particular circuit or gate behaves. The truth table shows the output from the circuit for all possible combinations of input states.

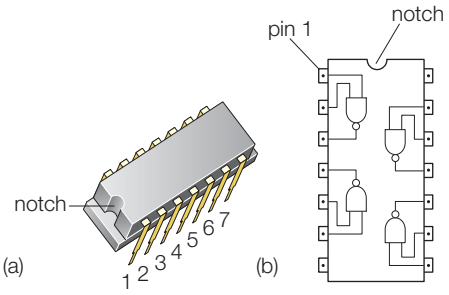


Figure 31.14 An integrated circuit (IC) consisting of four logic gates. (a) External view. (b) The 'pin' connections for the four gates.

The gate symbols used throughout this chapter are those used in the American system.

Controlling the input voltage (or logic state)

A potential divider arrangement (see page 402) may be used to provide the input voltage required for operation of a gate. Figure 31.15 shows a 6 V supply across two resistors R_1 and R_2. By Ohm's law, the voltage *drop* in a series circuit is proportional to resistance. Thus the voltages across R_1 and R_2 would be in the ratio of 100 Ω to 500 Ω or 1:5. The voltage drops across R_1 and R_2 are 1 V and 5 V, respectively. Thus, the terminal at A, which is connected to the gate input, would be at a potential of 6 V – 1 V = 5 V.

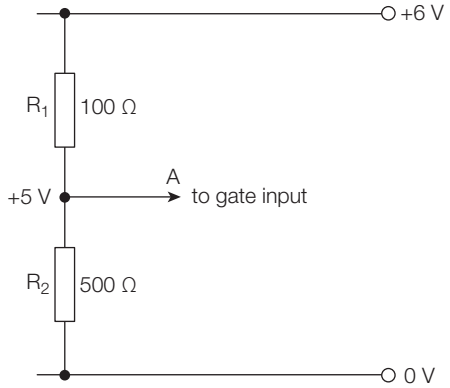


Figure 31.15 A potential divider arrangement. The voltage drop across resistor R_2 is given by $R_2/(R_1 + R_2) \times 6$ V, where R_1 and R_2 are the values of resistors R_1 and R_2.

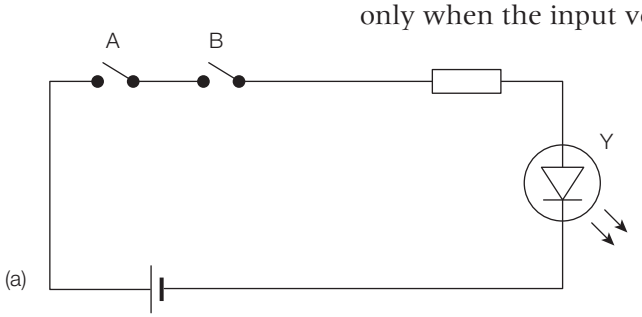
ITQ3

In the circuit shown in figure 31.15, if R_1 were 500 Ω and R_2 were 100 Ω, what would be the voltage to the gate input?

The AND gate

AND gate ›

Look at the mechanical switching arrangement in figure 31.16 (a). In this circuit, current would flow, and the light-emitting diode (LED) light up, only when switch A AND switch B are both closed. In the electronic **AND gate**, whose symbol is shown in figure 31.16 (b), the output voltage at Y will be high only when the input voltage at A is high AND the input voltage at B is high.



(c)

Input		Output
A	B	Y
0	0	0
1	0	0
0	1	0
1	1	1

Figure 31.16 The AND gate. (a) A mechanically switched AND gate. Both switch A AND switch B must be closed for there to be a current. (b) Symbol for electronic AND gate. (c) Truth table for an AND gate.

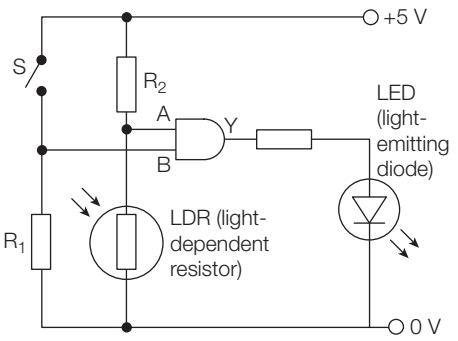


Figure 31.17 An AND gate operating an LED.

ITQ4

Suggest a practical use for a circuit such as shown in figure 31.17.

The truth table in figure 31.16 (c) shows the logic output of an **AND gate** for the possible combinations of logic inputs. For an output of logic state 1, input A AND input B must be at logic state 1.

The AND gate shown in figure 31.17 will turn on the LED only when it is both dark (input A is high, logic state 1) and the switch S is closed (input B is high, logic state 1). How does the circuit shown achieve this?

The light-dependent resistor (LDR) and resistor R_2 form a potential divider. The LDR resistance will only become high when the place is dark. If the resistance of R_2 is small compared with this high resistance of the LDR, the voltage drop across R_2 will be small. Thus, in darkness, a higher voltage will appear at A (logic state 1).

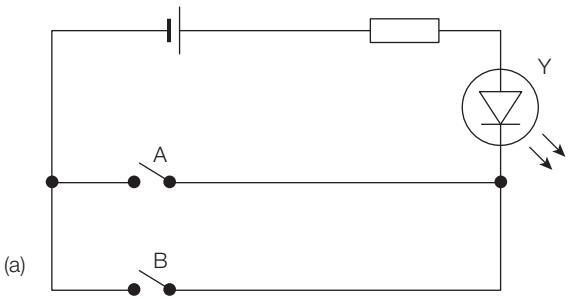
When switch S is open, there will be no voltage reaching B from the supply. Therefore the input voltage at B is 0 V (logic state 0). When S is closed, the voltage drop across S will be 0 V, and therefore B would be at a potential of 5 V (logic state 1)

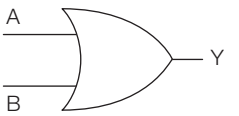
The high voltage inputs at A and B will give a 5 V output (logic state 1) at Y, which in turn causes the LED to light.

The OR gate

OR gate ›

In the **OR gate** shown in figure 31.18, a high voltage at the A input OR at the B input will result in a high voltage at the Y output.





(b)

(c)

Input		Output
A	B	Y
0	0	0
1	0	1
0	1	1
1	1	1

Figure 31.18 The OR gate. (a) Mechanically switched OR gate. (b) Symbol for electronic OR gate. (c) Truth table for OR gate.

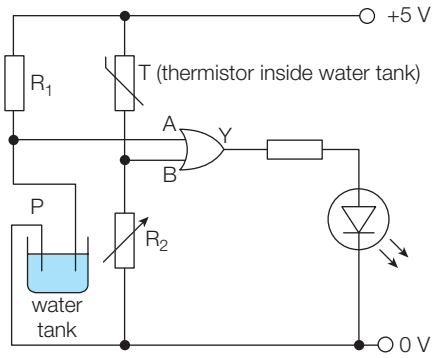


Figure 31.19 An OR gate operating an LED.

ITQ5

Why is a variable resistor placed in the thermistor circuit of figure 31.19?

In figure 31.19, the LED lights when either the water tank is empty OR the thermistor, T, is too hot. The resistance at T and resistor R_2 form a potential divider, as do those at the probes P and R_1.

When the tank is full, the gap at the probes P is bridged with water. If R_1 is much greater than the resistance of the water between the probes, most of the voltage drop will be across R_1, and so the voltage at A is low (logic 0). When the water level in the tank becomes very low, there will be a gap across the probes at P and therefore a high resistance. A high input voltage (logic 1) will therefore appear at A. This causes a large output voltage (logic 1, see the truth table in figure 31.19) at Y, which, in turn, causes the LED to light.

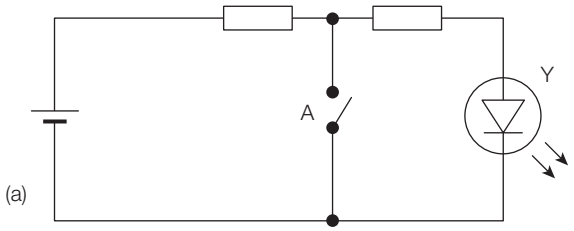
The thermistor T has a high resistance compared with resistor R_2. The voltage at B is therefore low (logic 0). If the thermistor gets hot, its resistance decreases and becomes small compared with R_2. Thus the voltage drop across the thermistor is now small, and there is large input voltage at B (logic 1). This also could cause the LED to light.

The circuit could therefore be used to given an alarm signal if the water level in a tank is too low OR if the water in the tanks gets too hot.

The NOT (inverter) gate

NOT gate ›

When a high voltage is applied to the input of a **NOT gate**, a low output is obtained (figure 31.20). When a low voltage is applied to the input, a high



Input	Output
A	Y
1	0
0	1

(c)

Figure 31.20 The NOT gate. (a) Mechanically switched NOT gate. Closing the switch at A (logic 1) forms a short-circuit and no current flows through the LED. (b) Symbol for electronic NOT gate (c) Truth table for NOT gate.

voltage output is obtained. A NOT gate is therefore sometimes called an **inverter**.

inverter ›

ITQ6

With reference to figure 31.21, complete the following using the words 'high', 'low', 'on' or 'off' in the spaces:
In the dark, the resistance of the LDR is ……. The voltage at A will be ……. The output voltage at Y will be ……. The LED will therefore be turned …….

Figure 31.21 shows how a NOT gate can be used to turn on a light automatically when it is dark. The potential divider is formed by the resistance of the LDR and the resistor R.

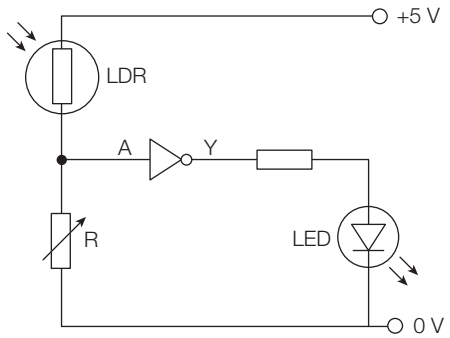
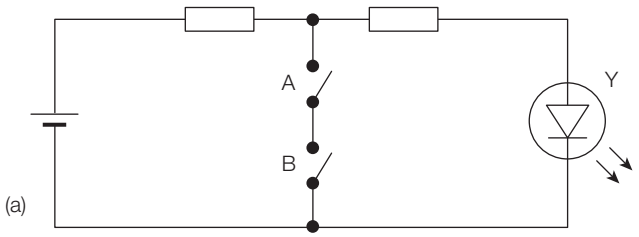


Figure 31.21 A NOT gate controlling an LED.

The NAND gate

NAND gate ›

The **NAND gate** (figure 31.22) functions in the *opposite* way to the AND gate. (NAND stands for NOT-AND.) Only when both inputs A and B are high does the output Y become low.

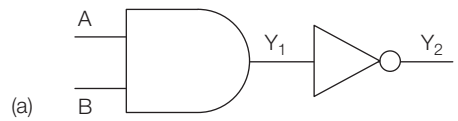


Input		Output
A	B	Y
0	0	1
1	0	1
0	1	1
1	1	0

(c)

Figure 31.22 The NAND gate. Any low input in such a gate results in a high output. (a) Mechanically switched NAND gate. (b) Symbol for electronic NAND gate. (c) Truth table for NAND gate.

A NAND gate could be formed by connecting the output of an AND gate to the input of a NOT gate (figure 31.23).

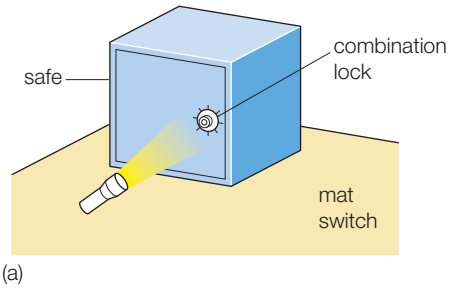


A	B	Y_1	Y_2
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	0

(b)

Figure 31.23 (a) A NAND gate can be formed by combining an AND gate and a NOT gate. (b) Truth table for AND gate followed by NOT gate.

The NAND gate circuit shown in figure 31.24 can be used to protect the safe. Since in the dark the LDR has a very high resistance compared with R_1, there will be a small voltage drop across R_1. The input at A will therefore be high (logic 1) when the room is dark. If the switch, S, is open (this happens when no one is on the mat), there will be no voltage drop across R_2 since there is no current in an open circuit. The input at B will therefore be high (logic 1). Since the inputs at Y will be low (logic 0) and the buzzer will not sound. (Recheck the truth table to verify this.) Thus, if the place is dark AND if no one is on the mat, the buzzer does NOT sound.



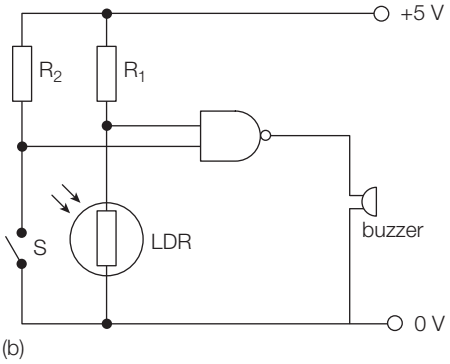


Figure 31.24 (a) A mat switch around a safe kept in a dark room. A light is needed to see the combination. (b) A NAND gate burglar alarm protecting the safe.

ITQ7

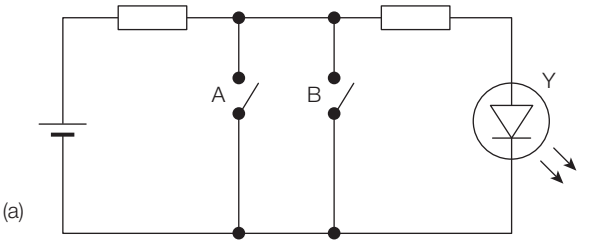
How can figure 31.24 be modified to serve as the same type of burglar alarm, but using an AND gate and one other gate instead of the NAND gate? Check your prediction using a truth table.

If someone steps on the mat, this closes switch S, and makes the input at B low. Any low input in a NAND gate results in a high output, which in turn causes the buzzer to sound. If a light was turned on, the resistance of the LDR would decrease. The input voltage at A would become low. Thus, again, the output would be high and cause the buzzer to sound. The alarm therefore sounds if someone steps on the mat or turns on a light.

The NOR gate

NOR gate ›

The **NOR gate** functions in the opposite way to an OR gate. When the input to A or B, or to both, is high, the output at Y is low (figure 31.25). The gate is therefore a NOT-OR gate. A NOR gate can be made by combining an OR gate with a NOT gate (figure 31.26).



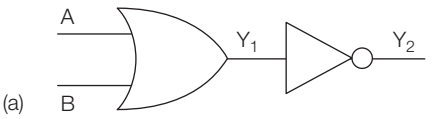
(c)

Input		Output
A	B	Y
0	0	1
1	0	0
0	1	0
1	1	0

Figure 31.25 NOR gate. (a) Mechanically switched NOR gate. (b) Symbol for electronic NOR gate. (c) Truth table for NOR gate.

ITQ8

In what way does the truth table for a NOR gate differ from the truth table for a NAND gate?



(b)

A	B	Y_1	Y_2
0	0	0	1
0	1	1	0
1	0	1	0
1	1	1	0

Figure 31.26 (a) A NOR gate can be formed by combining an OR gate and a NOT gate. (b) Truth table for OR gate followed by NOT gate.

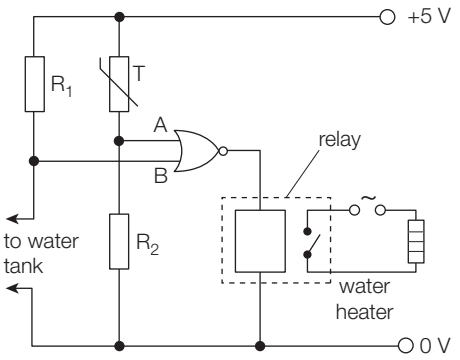


Figure 31.27 A NOR gate operating an automatic water heater.

Figure 31.27 shows a NOR gate being used to control a water heater. When the water is hot, the resistance of the thermistor T is low and there is a high input at A (logic 1). A high-level input (logic 1) occurs at B when the water level in the tank is low (i.e. resistance there is high). Either logic 1 input produces a logic 0 output, that is a low-voltage output, and the heater does not turn on. Thus, if either the tank is empty OR the water in the tank is hot, the heater will NOT turn on.

Combinations of gates

We have seen the effect of combining an AND gate with a NOT gate. We have also seen the result of combining an OR gate with a NOT gate. Figure 31.28 shows how an OR gate can be added to an alarm circuit to 'test' whether the alarm is working. Either a logic 1 output from the circuit or a logic 1 input to the OR gate will cause a logic 1 output to the alarm. Thus, when the 'test' button is pressed, input A becomes high, resulting in a high output, which can trigger the alarm.

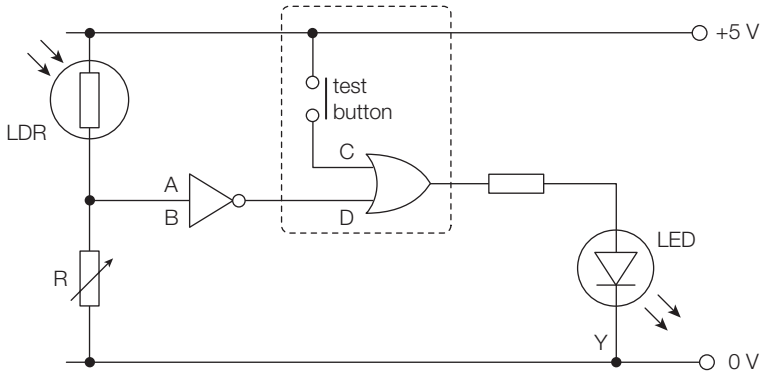


Figure 31.28 OR gate and 'test' switch added to a circuit.

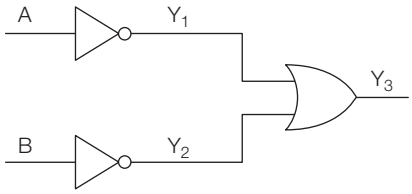
Practical activity 31.3

Combinations of gates

To what practical use can the circuit of figure 31.28 be put? Support your answer with a truth table.

ITQ9

Complete the truth table for the combination of gates shown in figure 31.29.



A	B	Y_1	Y_2	Y_3
0	0			
0	1			
1	0			
1	1			

Figure 31.29

Practical activity 31.4

Question 7 at the end of the chapter will give you further practice in arriving at truth tables from combinations of simple gates.

Combinations of gates

Using two NOT gates, one AND gate, two fixed resistors, an LDR, a thermistor, an LED with protective resistor, a suitable low voltage DC power supply and connecting wire, design a circuit that will light the LED if the room is dark and the temperature is high (as can happen in a storage room). Verify your design using a truth table.

Applications involving extensive electronic switching

In this chapter, we have explored simple circuit applications which employ electronic devices and switching. Many of the complex electronic devices today, such as computers, cell phones, medical equipment, automatic doors and robots, are just extensions of these building block ideas, but using millions or billions of switches within tiny chips.

The use of semiconductors has made switching achieve enormous frequencies. At one time, the switching was done by small currents in junction transistors. These currents can add up to much when we consider the billions of transistors being used in a single device. A consequence is wastage of energy. The development of field-effect-transistors (FET), however, has given rise to

switching using fields. Such miniscule currents are needed for switching that far less energy is required to operate the digital FET switches that make up complex circuits. One result of this is the increased talk-time available on cell-phones and increased computing time on computers per battery charge.

Another important development is the use of digital logic circuits (i.e. circuits involving on–off switching) to create smooth, analogue waveforms (such as sine waves). We experience the world in an analogue way – the sounds we hear and speak vary gradually in loudness and frequencies, the sights we see vary gradually in brightness. These gradual variations can now be achieved digitally through digital on–off electronic switching. For example, digitally produced sound can now be made indistinguishable from analogue sound to the human ear. It is usually far cheaper, uses less energy, and with better quality, to convert analogue sound to digital sound; then process, record and transmit the sound as digital signals; and finally receive, decode and reproduce the signals back to analogue sound.

Internet search term: digital electronics

Electronics applications continue to develop and transform marvellously the way we do things.

Chapter summary

- Mechanically operated switches may be operated by hand, by permanent magnets or by electromagnets.
- Electronic switches operate by means of voltages applied across p-n junctions of semiconductors.
- Electronic gate switches are operated by voltages that are in digital form (i.e. voltages that have specific values). In the binary system, these voltages are usually close to 0 volts (called logic 0) or approximately 5 volts (called logic 1).
- The input voltages to gates may be obtained from a d.c. voltage supply using potential dividers.
- A truth table gives the various possible outputs for different combinations of logic inputs to a gate or linked gates.
- Types of gate switches include AND, OR, NOT, NAND and NOR.
- Circuits involving gates and other semiconductor components are used to monitor conditions and control devices.
- Gates can be combined in several ways for practical use. Computers and digital musical instruments are only two of such applications.

Answers to ITQs

ITQ1 When the alarm sounds, if the door is quickly closed, the alarm will then stop. This is clearly a disadvantage.

ITQ2 Two terminals are connected to the relay coil and the other two to the on-off switch that lies inside the relay enclosure.

ITQ3 Voltage drop V across resistor R_1 is

$$V = \frac{R_1}{R_1 + R_2} \times 6\,\text{V} = \frac{500}{600} \times 6\,\text{V} = 5\,\text{V}$$

Voltage at input = 6 V – 5 V = 1 V.

ITQ4 As an automatic 'night-light' that goes on when the switch is on AND the place is dark.

ITQ5 By varying the value of R_2, we can set the temperature (being sensed by the thermistor) at which the alarm light must go on.

ITQ6 High; low; high; on.

ITQ7 By replacing the NAND gate with an AND and NOT gate in series (since an AND followed by a NOT is equivalent to a NAND).

ITQ8 In a NOR gate, only when both inputs are at logic 0 is the output at logic 1. In a NAND gate, only when both inputs are at logic 1 is the output at logic 0.

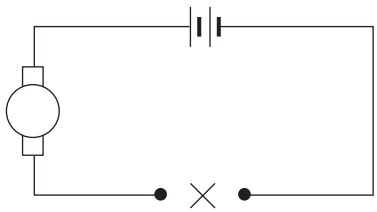
ITQ9

A	B	Y_1	Y_2	Y_3
0	0	1	1	1
0	1	1	0	1
1	0	0	1	1
1	1	0	0	0

Examination-style questions

Items 1–3

The diagram shows a circuit containing a direct current (d.c.) electric motor.



1 What device must be inserted at X so that the speed of the motor can be varied?

A capacitor
B diode
C fixed resistor
D rheostat

2 What device must be inserted at X to ensure that the motor spins in one direction only?

A diode
B light-dependent resistor (LDR)
C thermistor
D transformer

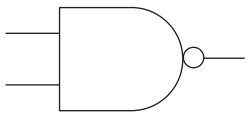
3 A thermistor, resting on the motor casing, is connected at X. If the motor heats up, what effect would be the MOST likely?

A The motor speeds up.
B The motor slows down.
C The motor comes to a stop.
D The motor reverses its spin.

4 Which gate is called an inverter?

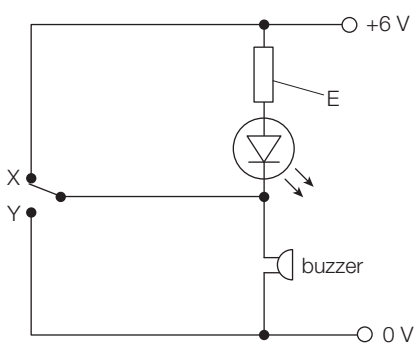
A AND
B OR
C NOR
D NOT

5 What is the name of the gate shown in the diagram?



A AND
B OR
C NAND
D NOR

6 The diagram shows a light-emitting diode (LED) and a buzzer connected to a 6.0 V power supply.

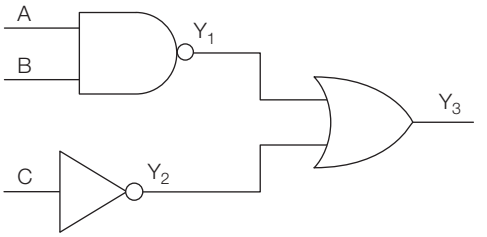


(i) For which position of the switch will the LED light? Explain your answer.

(ii) What is the name given to the component labelled E, and what is its function in the circuit?

(iii) If the current through the LED is not to exceed 20 mA, calculate the minimum value required for E.

7 (i) Complete the truth table for the combined gates shown in the diagram.



A	B	C	Y_1	Y_2	Y_3
0	0	0			
0	1	0			
0	0	1			
0	1	1			
1	0	0			
1	1	0			
1	0	1			
1	1	1			

(ii) Which combination(s) produce(s) a logic 1 output in Y_3?

(iii) Suggest a possible practical use of this type of circuit.

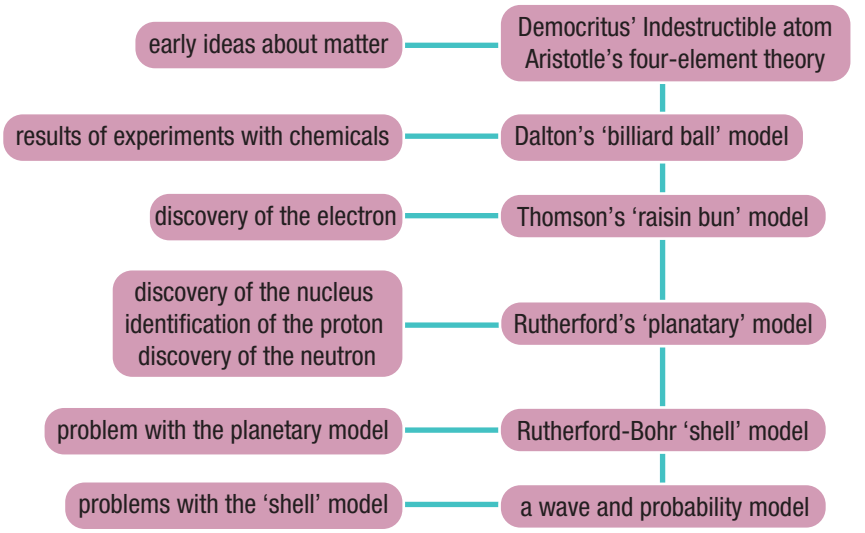
Section E:

The Atom

32 Models of the Atom

By the end of this chapter you should be able to:

- describe the work done by Dalton, Thomson, Rutherford, Chadwick and Bohr in establishing the modern view of the atom
- describe the Geiger–Marsden experiment, which established the nuclear structure of the atom
- recall that the nucleus contains positively charged particles (protons) and neutral particles (neutrons) of approximately equal mass
- state the approximate size of an atom



Early ideas

Over 2000 years ago, a group of Greek philosophers, led by Democritus (c. 470–380 BC), suggested that matter was made of tiny, indivisible particles, which they called 'atoms'. If one were able to break up a substance into tiny bits, they argued, one would reach a stage when the tiny piece could not be broken further. This indivisible piece was what they called the '**atom**'. Between these atoms that made up matter, there were empty spaces.

atom

Aristotle (c. 384–322), another famous Greek philosopher, disagreed with the atomic idea. He put forward instead that matter was continuous – like a

Did you know?
The word 'atom' comes from the Greek word *atomos*, which means 'indivisible'. Today, of course, we know that the 'atom' is divisible.

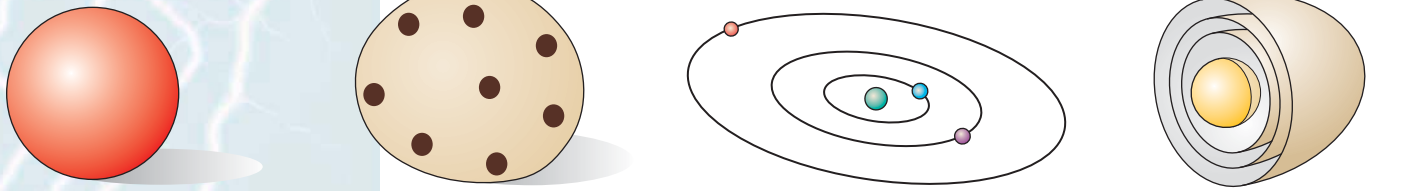


Figure 32.1 What is an atom like?

smooth jelly. Matter could be divided up into successively tinier pieces without limit. There were therefore no spaces in matter. Aristotle's ideas remained unchallenged for nearly 2000 years. Around the 13th century, however, an English monk, Roger Bacon, insisted that ideas about Nature obtained by reason must be tested by experiment. Bacon's ideas did not gain acceptance until about 300 years later. It was through insight, reason and experiment that we have the model of the atom as we know it today.

Dalton's model: the indestructible atom

John Dalton (1766–1844) was an Englishman who was both a teacher and a researcher. Dalton proposed an atomic model that was based firmly on experimental data. Dalton recognised from data available to him that, whenever two **elements** took part in a chemical reaction:

element

1 the total mass of the combining elements at the beginning of a reaction was the same as the mass of substance formed at the end of the reaction; and

2 their combining masses were in simple whole-number ratios.

For class (or group) discussion
As examples, 4 g of hydrogen combined with 32 g of oxygen to form 36 g of water and no other substance; 3 g of hydrogen combined with 24 g of oxygen to form 27 g of water and no other substance. Do these results support the two numbered statements above?

Internet search terms: chemistry and early atom

CHEMISTRY: the atom

Element
An element is a substance that cannot be broken down into anything simpler by chemical means. Hydrogen, mercury and copper are examples of elements. Water is *not* an element, since it can be broken down into simpler substances, namely hydrogen and oxygen.

To Dalton, since hydrogen and oxygen combined in a simple whole number ratio (1:8), this suggested that units of individual combining elements (i.e. atoms) had the same mass. Since no mass was lost during the combination reactions, this supported the view that the atoms must be indestructible.

Dalton's 'indestructible particle' theory of matter was similar to that of Democritus. Support for Aristotle's view diminished among the scientific community, since Aristotle's view was not backed up by experimental data.

Dalton's model became known as a 'billiard ball' model, since the atoms appeared to behave as billiard balls do: they remain whole when they strike one another (figure 32.2).

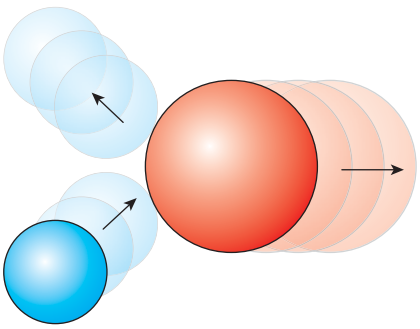


Figure 32.2 In Dalton's model of the atom, atoms interact without breaking up into smaller parts.

Thomson's model: a cloud with light, embedded particles

The discovery of cathode rays in the latter half of the 19th century helped to promote an even better understanding of the atom. Figure 32.3 shows an early cathode ray tube. The tube contained a gas at extremely low pressure. When a high voltage was applied across the electrodes, a pale green light appeared on the end of the glass tube near the anode.

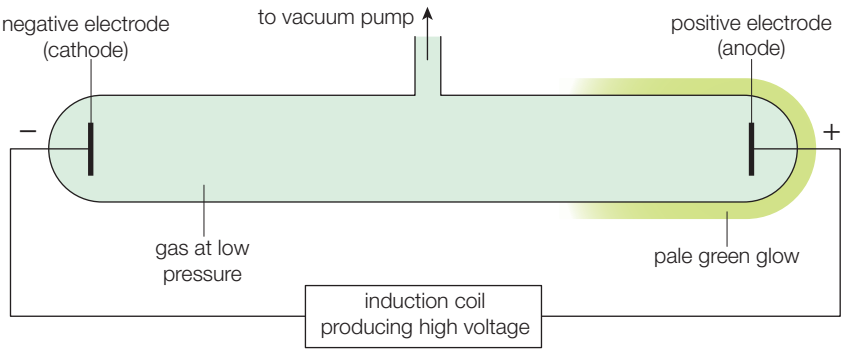


Figure 32.3 Discovery of cathode rays. The pale green light appearing at one end of the glass tube puzzled scientists.

Scientists were at first puzzled by the phenomenon of the glow. Later, they were able to verify that invisible rays were coming from the cathode. It was these rays that were colliding with the glass near the anode end, causing it to glow. The rays became known as '**cathode rays**'.

cathode rays ›

For class discussion
Can you think of a way of verifying whether the invisible rays causing the glow in figure 32.3 were coming from the cathode or the anode?

Discovery of the electron

Experiments with cathode rays indicated that these rays travelled in straight lines in a vacuum. In other experiments, a narrow beam of cathode rays was passed between two oppositely charged plates (see figure 32.4). The beam was deflected away from the negative plate, suggesting that the rays were negatively charged.

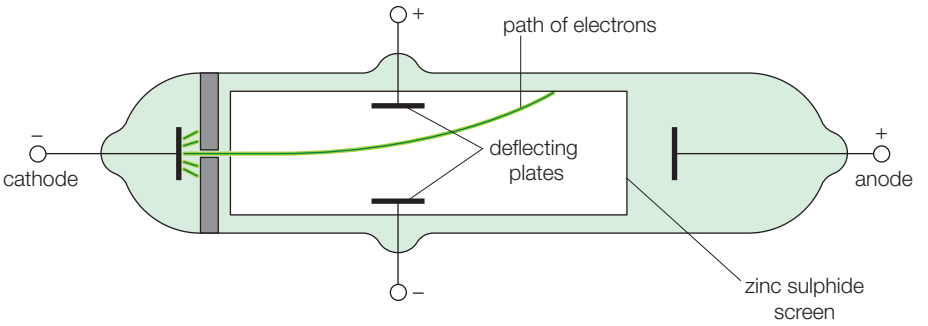


Figure 32.4 Cathode rays were deflected by an electrostatic field. The direction of the deflection suggested that the rays were negatively charged.

For class discussion
In figure 32.4, why does the path taken by the cathode rays suggest that the rays were negatively charged?

An Englishman, Sir Joseph J. Thomson (1856–1940), found from experiment that the 'cathode rays' behaved like very small, high-speed particles. Each particle had the same charge-to-mass ratio, no matter what metal was used for the cathode. Thomson's conclusions were based on experiments involving deflections of cathode rays by known electrostatic fields (fields between oppositely charged plates) in conjunction with known magnetic fields (fields produced by coils carrying currents). Thomson reasoned that these particles might therefore be present in all substances. Today we call these particles '**electrons**'.

electron ›

For years scientists had thought that all atoms might be composed of hydrogen atoms, since hydrogen was the simplest known unit of matter at the time. Thomson suggested that atoms themselves were composed of even smaller parts, the electron being one such small part.

Since the electrons were negatively charged, Thomson proposed that the neutral atoms from which they came must have had positive charge as well. He proposed a model of an atom in which electrons were distributed in a positive 'cloud' (figure 32.5). His model resembled a bun in which raisins were embedded. The raisins represented the electrons and the rest of the bun was the positive cloud that held the negatively charged electrons in place.

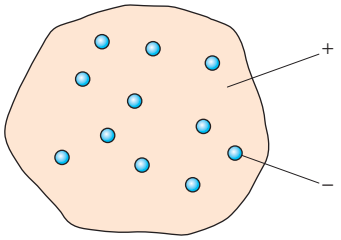


Figure 32.5 Thomson's model. The 'raisin bun' model resembled a 'mohanbog' served at Indian weddings. The pudding part represented the positive charge. The raisins represented the negative charge (electrons).

For class discussion
Can a guava fruit be used to illustrate Thomson's model of the atom? Why, or why not?

Rutherford's model: a nuclear, planetary model

An experimental physicist, Ernest Rutherford (1871–1937), designed a now famous experiment in 1911 to test Thomson's model.

The Geiger–Marsden experiment

Two of Rutherford's colleagues, Geiger and Marsden, aimed positively charged alpha (α) particles (alpha particles are discussed in chapter 34) towards a very thin sheet of gold foil. The experimental apparatus arrangement, simplified, is shown in figure 32.6. If Thomson's atomic model was correct (about electrons embedded in a cloud of positive charge), then most of the alpha particles would go straight through the foil, or be deflected by only tiny amounts (figure 32.7). This would be so since, according to the model, the electrons were very tiny and light, and there were no massive concentrations of positive charge within the atom.

It was as if a hail of bullets were fired at a thin sheet of cardboard, most going through, but one or two bouncing right back – suggesting that there must have been some tiny, but very heavy, things within the board.

For class discussion
1 Why do you think a very thin sheet was used?
2 What are two reasons for choosing gold in this experiment?

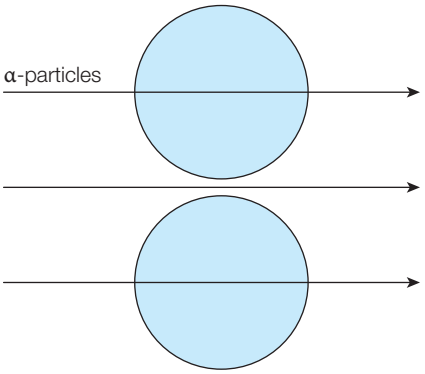


Figure 32.7 Predicted lack of scattering of alpha particles based on Thomson's model of the atom.

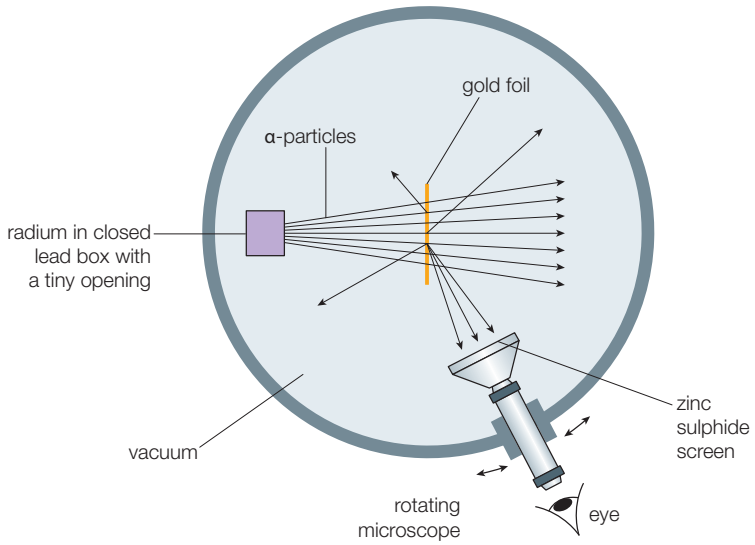


Figure 32.6 Geiger and Marsden's alpha particle scattering experiment.

Geiger and Marsden found that most of the alpha particles went straight through the foil, as Thomson's model predicted. However, a few were scattered through very large angles (figure 32.8), and some were even deflected directly back towards the source.

Rutherford concluded that most of the atom must therefore have been empty space. The positive part (which repelled the positively charged alpha particles) must have been heavy and concentrated in a very small volume. He called this tiny part the '**nucleus**' of the atom. Since the electrons had very little mass, then most of the mass of the atom must have been in the nucleus.

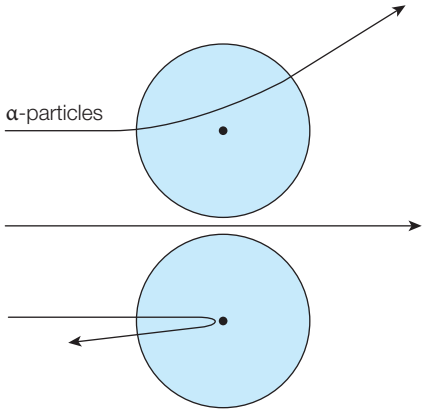


Figure 32.8 A nuclear model to account for the observed scattering of alpha particles.

nucleus ›

Rutherford estimated that the diameter of the nucleus was about 10^{-4} times the diameter of the atom.

For class discussion
The approximate diameter of an atom is 10^{-10} m. Using this value, estimate the diameter of the nucleus.

Did you know?
If the diameter of a nucleus were the size of a cricket ball, then the atom itself would have a diameter of about one kilometre. Most of matter is empty space.

The experiment had shown that matter was not continuous. In fact most of matter was empty space.

A nuclear, 'planetary' model

Rutherford suggested that the atom consisted of electrons orbiting a nucleus. The motion of an electron resembled that of a planet orbiting the Sun. A planet is kept in orbit around the Sun by a gravitational force of attraction, whereas electrons were kept in orbit by the force of attraction between the positively charged nucleus and the negatively charged electron. Rutherford's model of the atom was therefore sometimes called a 'planetary' model (figure 32.9).

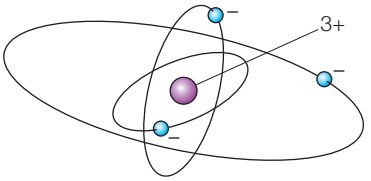


Figure 32.9 A nuclear, planetary model of the atom.

Structure of the nucleus

Rutherford knew that the hydrogen nucleus had the smallest charge of all known nuclei. He assumed that there was one electron orbiting the hydrogen nucleus. Since the hydrogen atom was neutral, he argued that the charge on the hydrogen nucleus must be equal and opposite to that on the electron. In 1920, Rutherford gave the name '**proton**' to the hydrogen nucleus.

proton ›

Experiments also showed that the charges on other nuclei were approximately whole-number multiples of the charge on the proton. All other nuclei, Rutherford reasoned, must therefore have a whole number of protons.

On the basis of Rutherford's model, a carbon nucleus, with 6 times the charge of a hydrogen nucleus, should have 6 protons, and thus 6 times the mass. Careful experiments revealed, however, that the mass of the carbon nucleus was 12, and not 6, times that of the hydrogen nucleus.

In order to explain this 'problem', Rutherford and his colleagues assumed that, within the nucleus, there must be another kind of particle with no charge. Rutherford proposed that particles with the same mass as a proton, but with no charge, were also present in nuclei. These particles became known as '**neutrons**'. The carbon nucleus must therefore have had 6 protons and 6 neutrons.

Neutron
The word 'neutron' is derived from 'neutral', which means neither positive nor negative. Note that the names of the three types of particles all end in the suffix 'on' – neutron, proton, electron.

neutron ›

Chadwick confirms the existence of the neutron

It was the English scientist, James Chadwick (1891–1974), who confirmed by experiment in 1932 that the neutron exists.

Chadwick, and other scientists, had been bombarding beryllium targets with alpha particles (figure 32.10). Very penetrating rays were produced, which were not affected by a magnetic or electrostatic field. Thus, these rays had no charge. Could they be electromagnetic waves? When the rays were directed at a block of paraffin (a compound containing hydrogen), protons were knocked out of the paraffin. On applying the laws of conservation of momentum and energy to the collisions, Chadwick concluded that the 'rays' with no charge behaved like particles having approximately the same mass as protons (see figures 32.10 and 32.11 and the margin note).

Chadwick had demonstrated by experiment that the particles were the neutrons whose existence Rutherford had predicted.

ITQ1
A nitrogen nucleus has a charge of 7 times that of a hydrogen nucleus. How many protons does the nitrogen nucleus contain?

ITQ2
A sodium nucleus has a charge 11 times that of a hydrogen nucleus, and a mass 23 times that of a hydrogen nucleus. How many protons and how many neutrons does the sodium nucleus have?

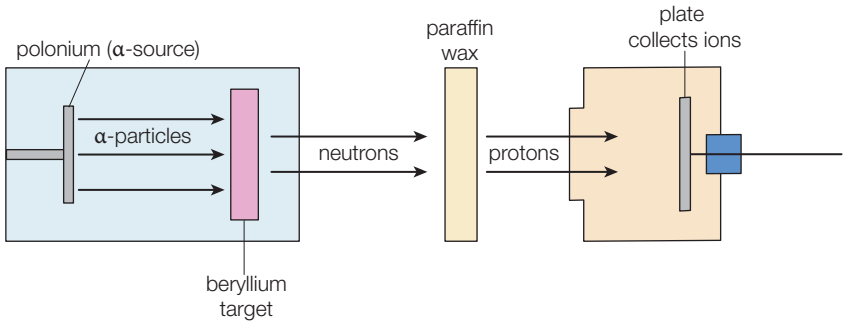


Figure 32.10 Chadwick's experiment to demonstrate that the neutron exists.

Internet search term: Newton's cradle animation

Chadwick's apparatus

An ionisation chamber was used for measuring the energies of the protons. These energies were assumed to come from the collisions of the protons with the uncharged 'rays' that were striking the paraffin block. By conservation of energy, the energy gained by a proton must have been equal to the energy lost by a particle from the uncharged 'rays' (much as when a moving pool or billiards ball strikes a stationary ball, the former comes to a stop and the latter gains its kinetic energy).

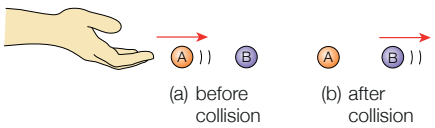


Figure 32.11

Colliding steel balls demonstration

A moving ball, A, collides with a stationary ball, B (figure 32.11). Ball A stops at the point of collision; ball B moves off with the original speed of ball A. This kind of behaviour occurs only if both balls have equal mass (and if the collision is elastic). Chadwick suggested that neutrons behaved like ball A and protons like ball B.

The nuclear model summarised

In summary, in the nuclear, 'planetary' model, the atom consisted of:

1 a tiny nucleus of large mass and containing positively charged particles (protons) and particles with zero charge and with mass approximately equal to that of protons (neutrons);

2 orbiting electrons, equal in number to the protons in the nucleus, and having a total charge equal, but opposite, to the total charge in the nucleus.

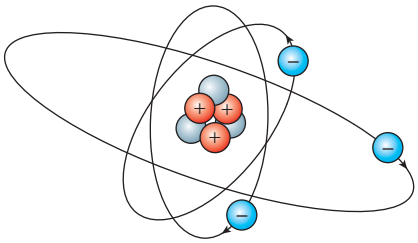


Figure 32.12 The nuclear, 'planetary' model for the lithium atom (note that the size of the nucleus has been greatly exaggerated).

CHEMISTRY: structure of the atom

An example of this nuclear, 'planetary' model is shown in figure 32.12.

The Bohr model: electrons arranged in 'shells'

By 1913, Niels Bohr (1885–1962), a Danish scientist, was working with Rutherford. They recognised that the Rutherford model had a serious flaw.

The problem of the collapsing atom

At the time, it was well known that a charge, moving in circle, would give off electromagnetic energy, since such a charge would be accelerating towards the centre of the circle. Thus, as an electron orbits a nucleus, it should give off electromagnetic energy. With the loss of energy, the negatively charged electron should eventually spiral towards the positively charged nucleus and come to a stop there. But this did not happen, since the electron and the nucleus in an atom were known to be far apart.

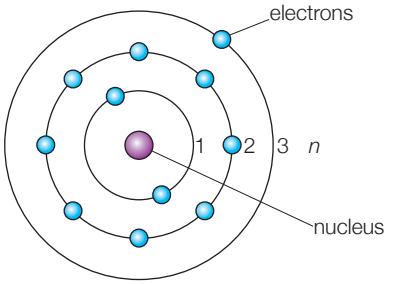


Figure 32.13 Bohr's 'shell' model of the sodium atom. The shells are spheres, but are drawn as flat circles for simplicity. The sodium atom has 11 electrons distributed as follows: 2, 8 and 1 in the $n = 1$, 2 and 3 shells, respectively.

Electron 'shells'

Bohr therefore proposed the following:

1 Electrons move around the nucleus in circular, 'allowed' orbits without losing energy.

2 Each orbit is associated with a particular energy level. Only when electrons possess this amount of energy do they occupy this orbit. Intermediate orbits are not possible.

An example of this model is shown in figure 32.13.

Bohr's allowed orbits became known as '**electron shells**'. Unlike the planetary model, each orbit (shell) could accommodate more than one electron.

electron shells ›

The energy levels of each shell were assigned 'quantum' numbers, $n = 1, 2, 3, \ldots$. A large n corresponded to a high **energy level**, and a small n to a low energy level. For atoms of the same element, all electrons in orbits with the same **quantum number** possessed the same amount of energy.

energy level ›

quantum number ›

The possible energy levels in the Bohr atom are like energy levels on a ladder (figure 32.14). One can achieve a stable potential energy state only if one stands on a rung of the ladder, not in between.

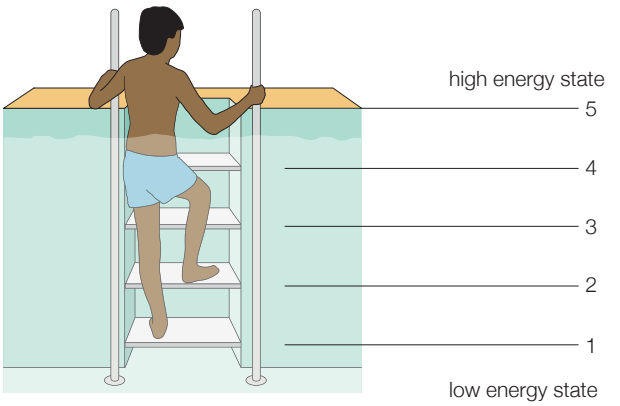


Figure 32.14

IT Q3

How many electrons can the $n = 2$ shell of an atom accommodate?

IT Q4

An atom of fluorine has 9 electrons. Using the shell model, draw a diagram to show the electron arrangement in such an atom.

Bohr further proposed that, for each n, the largest number of electrons that could occupy that energy level was $2n^2$. (Thus, for example, up to $2 \times 3^2 = 18$ electrons could occupy the $n = 3$ level.)

Success and failure of the Bohr model

The Bohr model of the atom successfully explained why only certain wavelengths of light were given off from a hydrogen lamp. Each allowed orbit of the electron in a hydrogen atom is associated with a specific energy level. An atom emits a fixed quantity of electromagnetic energy (called a **quantum**) when an electron moves from one specific energy level to a lower specific energy level. Figure 32.15 shows four such transitions to the $n = 2$ level. Each quantum of energy given out results in a particular wavelength of light being emitted. (This is because the energy E of an electromagnetic wave is related to the wavelength λ of the wave by the equation $E = hc/\lambda$, where h is a special constant and c is the speed of light.) Each wavelength also corresponds to a particular colour of light.

quantum ›

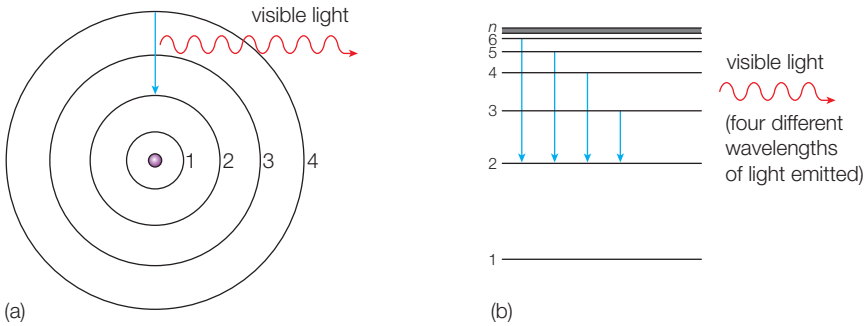


Figure 32.15 Energy level transitions in the hydrogen atom based on the Bohr model of the atom. (a) an Electron transition from the $n = 4$ shell to the $n = 2$ shell. Light of a particular wavelength is given off due to the transition. (b) Energy level diagram showing four transitions to the $n = 2$ energy level. The colours of light given off by five transitions are shown in figure 32.16 (a).

The Bohr atomic model could not, however, accurately account for wavelengths of light produced from elements other than hydrogen (i.e. elements with many-electron atoms). A better model of the atom was needed.

The Bohr model and atomic spectra

A spectrometer is an optical instrument for finding out what colours (wavelengths) are present in light that falls on it.

When substances other than hydrogen in the vapour state are given sufficient energy, they too give off characteristic colours of light. Thus, the light from a sodium vapour lamp looks yellow. Mercury vapour lamps give off a greenish white light. When the light from these lamps is viewed through a spectrometer, a unique set of colours (called *emission* lines) is seen for each vapour used in the lamp.

Figure 32.16 shows the spectra of coloured light obtained from a hydrogen lamp, a sodium vapour lamp, a mercury vapour lamp and the Sun. The Bohr model was able to account quite well for the **spectrum** obtained from the hydrogen lamp only.

spectrum ›

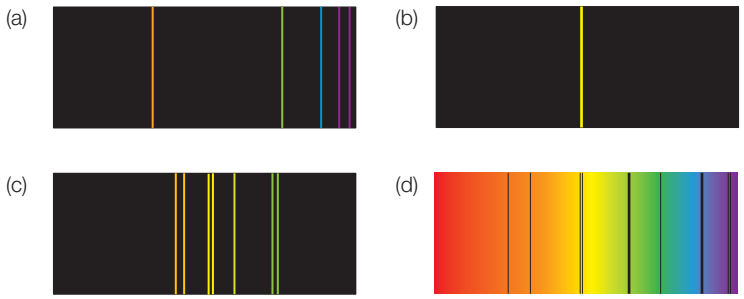


Figure 32.16 Visible spectra of light from: (a) hydrogen lamp, (b) sodium vapour lamp, (c) mercury vapour lamp and (d) the Sun. Note the prominent yellow colour characteristic of the light from a sodium lamp. The dark lines are absorption lines.

Absorption lines occur when emission lines of light from a given element pass through a vapour or gas of the same element. The atoms absorb each wavelength of emitted light, then re-emit them in all directions. As the lines are hard to see they are called 'absorption lines'.

Although the Bohr model had its limitations, it showed that there was a link between the spectrum emitted by an element and the energy level arrangement of the electrons in the atoms of that element. Thus, the model opened up an understanding of the outer structure of the atom. Scientists are now able to identify elements present in substances by their **emission spectra**. This development has proved to be valuable in analytical chemistry. In industry, spectral analysis enables monitoring of the composition of scrap iron used in the manufacture of steel. Helium was discovered in the Sun by analysing the Sun's spectrum.

emission spectrum ›

Did you know?

A forensic scientist can analyse the emission spectrum produced from a tiny flake of paint (e.g. after a hit-and-run car accident). The composition of the paint can therefore be determined and traced to the make and year of the car.

An understanding of spectra has also led to the development of lamps using elements such as krypton, xenon and neon. Some of these lamps are used with aesthetic effect in advertising signs. Light from sodium and mercury lamps is used widely for lighting of roadways and bridges.

A wave model of the atom

A model has been proposed, in which a wave rather than a particle approach is used to account for spectra not explainable by the Bohr model.

wave-probability model ›

A knowledge of the **wave-probability model** is not necessary at this level. The model has been introduced here only to show that atomic models are continually developing.

The following three assumptions were made in this model:

1 The electron is treated as having wave properties. In other words, the electron is thought to behave more like a wave than like a particle.

orbital ›

2 The position of electrons within the atom is treated on the basis of probability. Thus, electrons are found not in definite orbits but in **orbitals**. (An orbital is a volume of space in which it is very probable that an electron can be found.) Orbitals have specific shapes (figure 32.17), and this helps to explain the shapes of some molecules.

3 As in the Bohr model, quantum numbers are assigned to each orbital (energy state). But for each given n, there are n types of orbitals (designated s, p, d, f, …) n^2 orbitals in all. Each orbital can accommodate a maximum of 2 electrons.

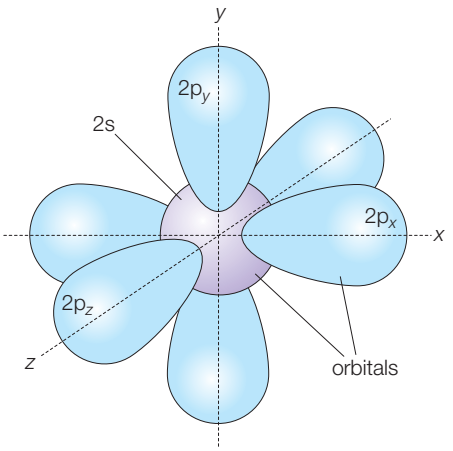


Figure 32.17 The shapes of some atomic orbitals in an atom.

Chapter summary

- According to Dalton's theory, matter was made up of indestructible particles (atoms). Atoms of the same element were alike. Atoms of different elements were different.
- J.J. Thomson is credited with discovering the electron in 1897. In Thomson's model of the atom, electrons were dispersed in a positive 'cloud'.
- The Rutherford/Geiger–Marsden alpha particle scattering experiment showed that most of an atom was empty space; most of the mass of an atom was contained in a tiny nucleus, which was positively charged.
- Rutherford also identified, within the nucleus, positively charged particles called protons, and predicted the presence of particles called neutrons, which had zero charge.
- Chadwick demonstrated by experiment that neutrons exist.
- The problem posed by Rutherford's model was that an orbiting electron would give off electromagnetic energy, and, as a result, would soon crash into the nucleus.
- Bohr proposed a model in which there were 'allowed' orbits or 'shells' in which electrons could move without giving off energy. An electron gained or lost a specific quantum of energy as it moved from one given shell (energy level) to another shell.
- Each shell (energy level) was assigned a quantum number, $n = 1, 2, 3, \ldots$. An atom would emit light of a particular wavelength when an electron moves from a given energy level to a given, lower energy level (intermediate orbits were not allowed).
- One problem posed by the Bohr model was that it could not account for wavelengths of light emitted from atoms with several electrons.
- A wave-probability model of the atom has achieved great success in accounting for energy levels of electrons within the atom. In this model, the electron is treated as a wave and not as a particle. The position of an electron in the atom is described using probability; orbitals, rather than orbits, describe the electron's position.
- The diameter of an atom is about 10^{-10} m.

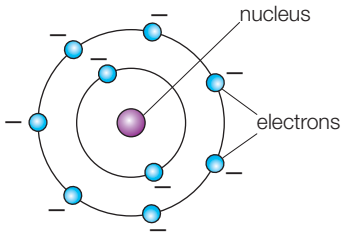
Answers to ITQs

ITQ1 7 protons.

ITQ2 11 protons, 12 neutrons.

ITQ3 $2 \times 2^2 = 8$

ITQ4



Examination-style questions

1 Name the person who is generally credited with developing each of the following models of the atom:

(i) billiard ball model
(ii) raisin-bun model
(iii) nuclear, planetary model
(iv) shell model.

2 One the basis of the shell model, what is the maximum number of electrons allowed in each of the first three shells (n = 1 2, 3) of an atom?

3 Using the shell model, draw diagrams to show the electron arrangement in an atom that has:

(i) 5 electrons
(ii) 22 electrons.

4 State the assumption that Bohr made to overcome the problem (in the Rutherford model) where an orbiting electron would eventually collapse in to the nucleus.

5 State *one* success and *one* failure of the Bohr model.

6 State *two* ways in which the modern view of electron behaviour differs from Bohr's view.

7 Particles leaving the cathode in a cathode ray tube are:

A atoms
B electrons
C neutrons
D protons

8 The Rutherford/Geiger–Marsden alpha particle scattering experiment demonstrated the existence of the:

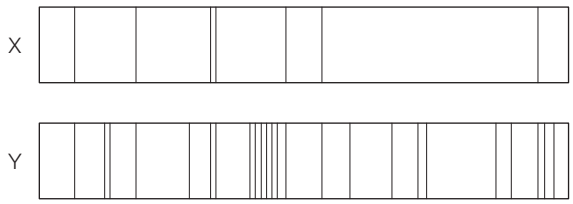
A electron
B neutron
C nucleus
D orbital

9 State the approximate diameter of an atom.

10 Which of the following are found in the nucleus of an atom?

A electrons only
B electrons and protons
C protons and neutrons
D protons only

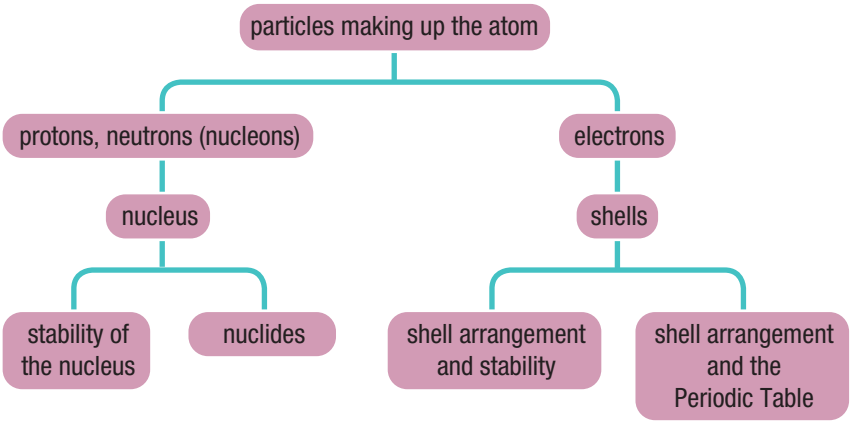
11 The emission spectrum of an element, X, found at a scene of crime, is shown in the diagram. The emission spectrum from a mixture, Y, found on a suspect's clothing, is also shown in the diagram. Could the element, X, have been present in the mixture, Y? Explain your answer.



33 Structure of the Atom

By the end of this chapter you should be able to:

- sketch the structure of simple atoms
- compare the charge and mass of the electron with the charge and mass of the proton
- explain why an atom is normally neutral and stable
- represent diagrammatically the structure of simple atoms
- relate the shell model of the atom to the Periodic Table
- recall that an atom of any element in the Periodic Table has one more proton than an atom of the element before it
- recall and use the relationship $A = Z + N$
- use the standard notation for representing a nuclide, e.g. $^{12}_{6}C$
- explain what is meant by the term 'isotope'



Particles that make up the atom

Bohr's 'shell' model of the atom has its limitations (see chapter 32). However, because it is fairly easy to visualise, this model is used in the discussions in this chapter.

In chapter 32, we saw that the atom consists of three types of particles: electrons, protons and neutrons. According to the Bohr model, electrons are negatively charged and revolve in allowed orbits called 'shells' around the nucleus. The nucleus itself contains protons, positively charged particles, as well as neutrons, particles with zero charge.

Charge on the electron, proton and neutron

In 1915, Robert A. Millikan, an American scientist, succeeded in measuring the charge on an electron. He obtained a value of approximately 1.6×10^{-19} C (coulombs). He also found that all other charges appeared to be integral (whole-number) multiples of this charge. The magnitude of this charge has been designated e, the **elementary charge**. Hence the charge, Q, on any object is given by the equation

elementary charge, e

$$Q = \pm ne$$

electron ›

where n is a whole number. The ± sign tells whether the charge is positive or negative. Since the **electron** is negatively charged, the charge on an electron can be stated as $-e$ or -1.6×10^{-19} C. Chemists commonly represent the electron charge simply as –1 and the charge on the proton as +1.

proton ›

The simplest atom, the hydrogen atom (see figure 33.1), consists of one **proton** and one electron. Since the hydrogen atom is neutral, it means that the charge on the proton is $+ 1.6 \times 10^{-19}$ C, or $+e$.

neutron ›

James Chadwick had shown in 1932, by experiment, that the **neutron**, a particle with approximately the same mass as a proton, had zero charge.

Table 33.1 summarises the charges present on an electron, proton and neutron, respectively, in terms of e.

Table 33.1 Charges on particles.

Particle	Charge
electron	$-e$
proton	$+e$
neutron	0

ITQ1

How many electrons are there in a charge of magnitude 1 C? (Hint: Use the equation $Q = \pm ne$.)

Mass of the electron, proton and neutron

charge-to-mass ratio ›

Thomson's experiment in 1898 had yielded a **charge-to-mass ratio** (e/m) of approximately 2×10^{11} C kg^{-1} for the electron. (The accepted value today is 1.76×10^{11} C kg^{-1}.) Using today's accepted e/m value, as well as the value of e obtained from Millikan's experiment, we can estimate the mass of an electron.

Group discussion: mass of an electron
How can we estimate the mass of an electron from the accepted values of e/m and e above? What is the value of this mass?

unified atomic mass unit, u ›

The accepted values for the masses of the three types of particles are given in table 33.2. Note that the masses are given in kg as well as in **unified atomic mass units**, **u**, where $1\text{ u} = 1.6605 \times 10^{-27}$ kg. The data in the table show that the mass of an electron is about 1/1840 that of a proton. A proton has approximately the same mass as neutron.

Unified atomic mass unit, u
1 u = $^1/_{12}$ the mass of a carbon-12 atom.
Expressed in kg, $1\text{ u} = 1.6605 \times 10^{-27}$ kg.

Table 33.2 Masses of particles.

Particle	Mass/kg	Mass/u	Mass/u (to nearest whole number)
electron	9.1094×10^{-31}	0.000 548 58	0
proton	1.67262×10^{-27}	1.007 276	1
neutron	1.67493×10^{-27}	1.008 665	1

Stability of the atom

nucleon ›

There are two types of particles within the nucleus of the atom – protons and neutrons. These particles are called **nucleons**. (In the diagrams, the sizes of the nuclei are drawn exaggerated to show the protons and neutrons present. The nucleus, in reality, is very tiny relative to the size of an atom.) Orbiting the nucleus are the electrons.

Stability of the nucleus

The simplest atom, the hydrogen atom, is shown in figure 33.1. The nucleus of the hydrogen atom consists of one proton, around which an electron orbits.

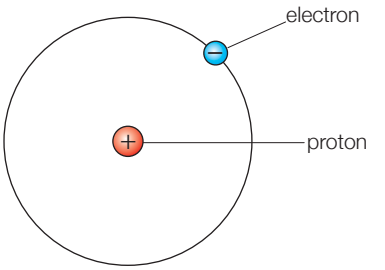


Figure 33.1 A hydrogen atom.

The nucleus of the next simplest atom, the helium atom, contains two protons (figure 33.2). Since protons are positively charged, and like charges repel, we would expect such a nucleus to fly apart. The electrostatic force of the repulsion between the two charges is indeed very great. So, what is it that keeps the nucleus from breaking up?

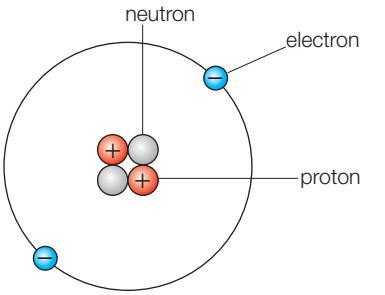


Figure 33.2 A helium atom.

The answer must lie in the presence of the neutrons. Although neutrons have zero charge, it would appear that they help to hold the positive charges together inside the nucleus. Thus, the neutrons contribute to the stability of the nucleus by helping to keep the protons together.

strong nuclear force ›

The force that keeps these particles together is called the **strong nuclear force**. The strong nuclear force acts only within the nucleus and on protons and neutrons alike.

The nucleus of the helium atom shown in figure 33.2 contains 2 protons and 2 neutrons. The nuclei of atoms of different elements are made up of different numbers of protons and neutrons. An **element** is distinguished from another by the number of protons present in the nuclei of each of its atoms. The number of protons present in the nucleus of an atom is called the **atomic number** (or **proton number**), Z, of the atom. The number of neutrons present within a nucleus is called the neutron number, N.

element ›

atomic number ›

proton number ›

stable ›

An equal number of protons (Z) and neutrons (N) results in **stable** nuclei up to about $Z = 30$. Nuclei with larger numbers of protons require more neutrons than protons in order to remain stable because the total repulsive force among the protons becomes excessively large. However, the presence of too many neutrons in a nucleus can also contribute to instability. This is due to the raising of the overall energy level of the nucleus into an unstable state as the number of neutrons increases.

Group activity: nuclear stability graph
In view of the above paragraph, sketch what a graph of neutron number (N) against proton number (Z) would look like for stable nuclei for Z up to about 80.

ITQ2

Is the nucleus of the element astatine (atomic number 85) likely to be stable or unstable?

Stable energy states among the orbiting electrons

It is now believed that, second only to nuclear forces, electrostatic forces are the most powerful forces present in Nature. Thus, positive and negative charges attract each other, resulting in a net zero charge on most objects. It is therefore not surprising that atoms in their normal state are electrically **neutral**. An atom whose nucleus has 8 protons, for example, has, in its normal state, 8 electrons orbiting the nucleus, giving rise to a net charge of zero. We say that such an atom is electrically neutral.

neutral ›

Electrons are kept in orbit by the electrostatic forces of attraction between the positive charges in the nucleus and the negatively charged electrons. However, there are repulsive electrostatic forces among the orbiting electrons themselves. This attraction by the nucleus and repulsion among electrons results in electrons settling in certain 'allowed' stable energy states, designated by quantum numbers, $n = 1, 2, 3, \ldots$.

According to Bohr model, the possible energy states correspond to orbits of differing radii or 'shells' around the nucleus. The maximum number of electrons that could be accommodated in a shell numbered n, according to this model, is $2n^2$. The innermost orbit is represented by the quantum number $n = 1$, and can accommodate 2 electrons. The next orbit, for which $n = 2$, can accommodate 8 electrons.

IT Q3

According to this pattern, what is the maximum number of electrons that can occupy the third orbit ($n = 3$)?

An atom tends to achieve a stable energy state when its outermost shell contains the maximum allowable number of electrons

The Periodic Table

Groups

Periods	I	II											III	IV	V	VI	VII	VIII
1	1 H 1.00																1 (H) 1.00	2 He 4.00
2	3 Li 6.94	4 Be 9.01											5 B 10.8	6 C 12.0	7 N 14.0	8 O 16.0	9 F 19.0	10 Ne 20.2
3	11 Na 23.0	12 Mg 24.3											13 Al 27.0	14 Si 28.1	15 P 31.0	16 S 32.0	17 Cl 35.4	18 Ar 39.9
4	19 K 39.1	20 Ca 40.1	21 Sc 45.0	22 Ti 47.9	23 V 51.0	24 Cr 52.0	25 Mn 54.9	26 Fe 55.9	27 Co 58.9	28 Ni 58.7	29 Cu 63.5	30 Zn 65.4	31 Ga 69.7	32 Ge 72.6	33 As 74.9	34 Se 79.0	35 Br 79.9	36 Kr 83.8

Figure 33.3 Periodic Table showing the first 36 elements and their respective atomic numbers and relative atomic masses.

Periodic Table ›

The elements have been arranged in an atomic number sequence to form a **Periodic Table** (figure 33.3). In this sequence, an atom of an element has one more proton than an atom of the element preceding it. The elements are arranged in horizontal rows, called **Periods**, and vertical columns, called **Groups**. In order to relate the shell model to this table, we need to focus on three aspects of the shell model:

Period Group ›

1 Electrons are arranged in shells numbered 1, 2, 3, … n.

2 Each shell can accommodate up to $2n^2$ electrons.

3 Electrons occupy the innermost shells of an atom first, up to the maximum capacity of each shell.

CHEMISTRY: Periodic Table

Figure 33.4 shows shells (numbered 1, 2, 3) and their maximum electron capacities in an atom. A fluorine (F) atom with atomic number 9 will have 9 protons, and hence 9 electrons. The innermost shell ($n = 1$) will have 2 electrons and be full. The remaining 7 electrons will be in the next shell ($n = 2$), since the second shell can accommodate up to 8 electrons. The electron arrangement in this atom is written 2, 7. Similarly, the electron arrangement in the magnesium (Mg) atom (atomic number 12) is 2, 8, 2.

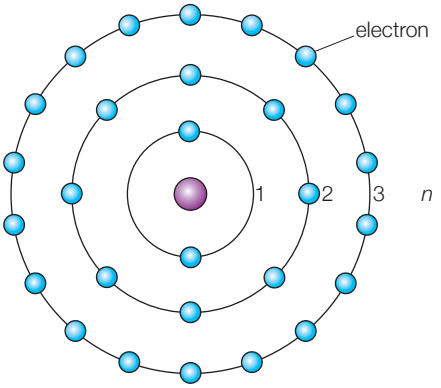


Figure 33.4 Shell model of the atom showing the maximum number of electrons accommodated in the 1st to the 3rd shells.

Periods

Atoms whose outermost electrons are in the same shell are placed in the same Period. Thus hydrogen and helium are in Period 1. The elements lithium to neon are placed in Period 2, and sodium to argon in Period 2. The electron arrangements in atoms of some elements from Periods 1 to 4 are shown in table 33.3.

Element symbol	*Z*	Electron arrangement
H	1	1
He	2	2
Li	3	2, 1
Be	4	2, 2
F	9	2, 7
Ne	10	2, 8
Na	11	2, 8, 1
Mg	12	2, 8, 2
Al	13	2, 8, 3
Cl	17	2, 8, 7
Ar	18	2, 8, 8
K	19	2, 8, 8, 1
Ca	20	2, 8, 8, 2
Sc	21	2, 8, 9, 2

Table 33.3 Electron arrangement within shells of atoms of selected elements. (Scandium (Sc) does exactly fit the pattern, but the reason for this is outside your course.)

Groups ›
outermost shell ›

Each **Group** consists of atoms that have the same numbers of electrons in the **outermost shell**. So in Group I, the atoms of hydrogen, lithium, sodium and potassium all have 1 outer electron. These atoms achieve a stable state by losing this outermost electron. Group VII (nowadays numbered as Group 0) consists of the halogens. The outermost shell of fluorine (F) and chlorine (Cl) each has 7 electrons. These atoms achieve a stable state by gaining 1 electron. Group VIII consists of elements like helium (He) and neon (Ne) whose outermost shells are full. These are the inert elements. It is the number of electrons in the outermost shell of an atom that contributes significantly to that element's chemical character. Thus, sodium and potassium (Group I elements) react violently with water to form alkalis. Fluorine and chlorine (Group VII elements) form acids with hydrogen.

Elements in Groups I to III are classified as metals. During chemical reactions, these tend to achieve stability by giving up their outermost electrons, leaving the next inner shell full. Those in Groups V to VII are classified as non-metals. In chemical reactions, these tend to receive electrons into their unfilled outermost shells. Elements placed in Group IV have some very special properties due to mutual sharing (silicon and germanium, for example are

CHAPTER 30

semiconductors and behave electrically quite differently from conductors – see chapter 30; carbon has the ability to form very long molecular chains).

Within each Group, comparing Period 2 with Period 3, for example, we see that chemical behaviour patterns in Period 2 are repeated in Period 3. Hence the term 'Period' is applied to each row of elements in the table because the behaviours within each Period are, more or less, periodic, i.e. repeated in each succeeding row.

Group discussion: Periodic Table (figure 33.3)

1 Which *two* elements should behave chemically like magnesium (Mg)? Explain in terms of electrons.
2 How many metals are shown in figure 33.3?
3 Would elements in Group VIII (Group 0) conduct electricity? Explain your answer.

Limitations of the shell model

A closer look at the Periodic Table (figure 33.3) would reveal some discrepancies with the shell model. Potassium (K), for instance, is placed in Period 4. With an atomic number of 19, and hence 19 electrons in the atom, we would expect, according to the shell mode, an electron arrangement of 2, 8, 9. This is because the third shell ($n = 3$) can accommodate up to 18 electrons. Potassium is placed in the Period 4 instead, since its electron configuration is 2, 8, 8, 1, i.e., its outermost electron is actually in the 4th shell.

Remember that the shell model of the atom is only a model. The model does not work well for atoms with many electrons.

Again, an atom of the inert gas argon (Ar) has 18 electrons. The electron arrangement in the argon atom should therefore be 2, 8, 8. The third shell is not full (it can contain up to 18 electrons). Yet argon is very unreactive, i.e. it behaves as though the outer (i.e. 3rd) shell was full.

The wave-probability model of the atom, mentioned in chapter 32, overcomes these and other discrepancies. However, knowledge of the wave-probability model of the atom is not required at this level.

Bonding between atoms

Metallic bonding

CHAPTER 24

In chapter 24, in the section 'Electron theory and conduction', we mentioned that metals are good conductors of electricity. This is because the outermost shell of an atom of a metal has few electrons (typically 1 to 3). These electrons are relatively easy to dislodge from their parent atoms, leaving ions formed, with their outer shells full; the electrons are not bound to individual atoms. The result is a vibrating lattice of ions throughout which electrons meander. The electrostatic forces, however, between the positively charged lattice of ions and the electrons are very great, holding them all together in a bonding that is called **metallic bonding**.

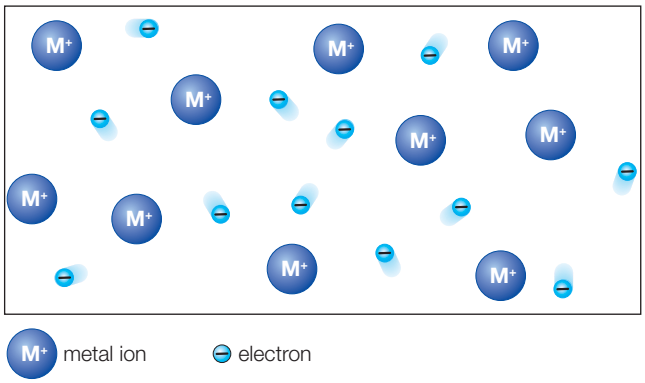


Figure 33.5 Random movement of outermost electrons in an ionic lattice in a metal.

Internet search terms: metallic bonding – YouTube tutorial

CHEMISTRY: bonding

metallic bonding ›

Other types of bonding

electrovalent (ionic) bonds ›

Atoms of metals (Groups I–III) typically form **electrovalent** (**ionic**) **bonds** with atoms of halogens (Group VIII). Figure 33.6 shows that an atom of the metal lithium gives up an electron to the fluorine atom, forming such a bond. In the process, atoms of both elements have their outer shells full – a stable state.

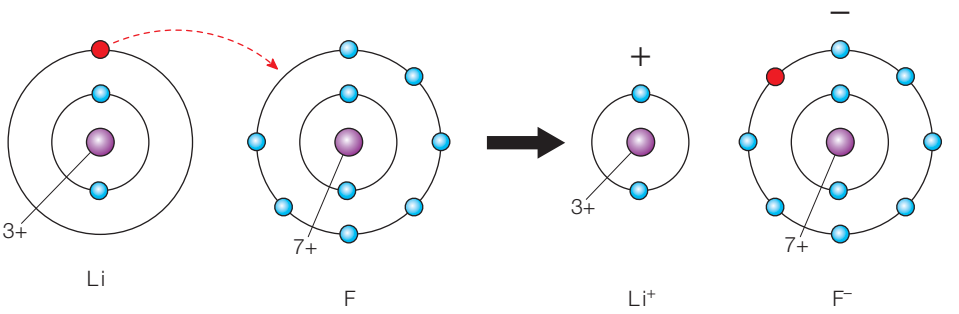


Figure 33.6 Formation of stable states (outermost shells filled) by electron transfer from a lithium atom to a fluorine atom.

covalent bonds ›

Other atoms can form **covalent bonds** through mutual sharing of electrons. Elements in Group IV tend to form covalent bonds with others readily. In covalent bonding, mutual sharing of electrons among atoms results in the outermost shells being filled to capacity – a very stable state, again (see figure 33.7). Also, in chapter 30, in the section on 'Semiconductor properties', covalent sharing was illustrated in the cases of silicon and germanium. Although a study of covalent bonding is not required at this level, the internet search words below will guide you to the unique property of carbon atoms in forming large chains, through covalent bonding, resulting in the formation of complex molecules, such as DNA, which form the very basis of life.

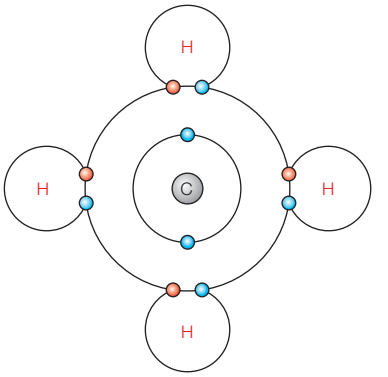


Figure 33.7 Achieving stability by mutual sharing of electrons. Both the carbon and the hydrogen atoms in this methane molecule have their outermost shells full by mutual sharing of electrons.

ITQ4

How do you think helium was discovered in the Sun? (Hint: Check 'The Bohr model and atomic spectra', chapter 32.)

If the outer shell of an atom is already full, then the atom will be in a stable state. It is unlikely that such an atom will take part in a chemical reaction. Hence, helium and neon (figure 33.8) are very unreactive elements. This is because the outermost shells of their atoms are full. In fact, helium was first discovered in the Sun, not on Earth, because it is so unreactive, chemically.

Internet search terms: covalent bonds/cell biology

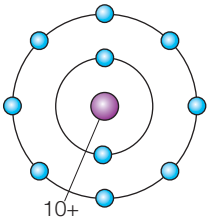


Figure 33.8 In a neon atom, the outermost shell, having 8 electrons, is full. This is a very stable state. Hence, neon is very unreactive.

Describing nuclides

nuclide ›

Nuclei of a given species (i.e. that have the same number of protons) are called **nuclides**. Three numbers are used in describing a particular nuclide – the atomic number, the neutron number and the mass number.

atomic number | proton number ›
neutron number ›
mass number | nucleon number ›

The **atomic number** (or **proton number**), Z, is the number of protons in the nucleus. The **neutron number**, N, is the number of neutrons in the nucleus. The **mass number** (or **nucleon number**), A, is the number of nucleons (proton and neutrons) present in the nucleus. In the lithium nucleus

Shortened form of representing a nuclide

Since the atomic number of each element is known, only the mass number is sometimes used to specify a particular nuclide. Thus, the atomic number of carbon is 6. The nuclide is sometimes written as ^{14}C and is referred to simply as carbon-14. Such a nuclide would have 6 protons and 8 neutrons.

shown diagrammatically in figure 33.9, $Z = 3$, $N = 4$ and $A = 7$. These numbers agree with the general formula

$$A = Z + N \qquad (1)$$

In a neutral atom, the number of protons within the nucleus is equal to the number of electrons orbiting the nucleus. It is the number (and arrangement) of the electrons in an atom that gives each element its unique chemical properties. Thus, we can identify an element by its atomic number.

A nuclide of an element is usually specified using the chemical symbol, X, of the element preceded by the mass number, A, and the atomic number, Z, as shown below:

$${}^{A}_{Z}\mathrm{X}$$

Thus the nuclide of lithium shown in figure 33.9 is written ${}^{7}_{3}\mathrm{Li}$. For this nuclide, the mass number $A = 7$ and the atomic number $Z = 3$. The neutron number, N, is calculated using equation (1). Thus, $N = A - Z = 7 - 3 = 4$.

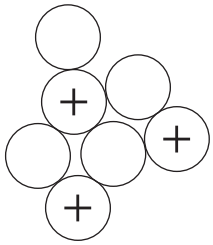


Figure 33.9 A lithium-7 nuclide. We specify this nuclide as ${}^{7}_{3}\mathrm{Li}$.

IT Q5

A chlorine nucleus contains 17 protons and 18 neutrons. Write down the symbol representing this nuclide. What is the mass number of this nuclide?

Isotopes

isotope ›

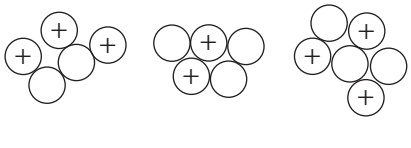
Elements that have the same atomic number (number of protons) but different mass numbers (number of nucleons) are called **isotopes**. Thus, carbon-12 and carbon-14 are isotopes of carbon. Each isotope of carbon has the same number of protons (6), and therefore the same number of electrons. Since it is the number of electrons that contributes most significantly to an element's chemical properties, it follows that both carbon-14 and carbon-12 will show identical chemical behaviour. Thus, isotopes are chemically indistinguishable from each other. Their nuclides differ from each other only in the number of neutrons (and hence in the total mass numbers).

For more on beta particles and radioactive isotopes see chapter 34.

Some elements have several isotopes. We have just pointed out two from carbon. The nuclides of some of these isotopes are radioactive. Thus, the nuclei of carbon-14 atoms give off beta particles (carbon-12 atoms, however, are stable).

IT Q6

Which two of the nuclides, A, B and C, represented diagrammatically in figure 33.10, belong to isotopes of an element, X? Write down the symbols representing these nuclides in terms of X.



A B C

Figure 33.10

Chapter summary

- The magnitude of the elementary charge $e = 1.6 \times 10^{-19}$ C and is the size of the charge on an electron. The charge, Q, on any object is believed to be a whole-number multiple, n, of the elementary charge. Thus $Q = \pm ne$.
- The unified atomic mass unit u = 1.6605×10^{-27} kg.
- An atom is made up of three types of particles: electrons, protons and neutrons. The particles have the following characteristics:

Particle	Charge/e	Approx. mass/u
electron	–1	1/1840
proton	+1	1
neutron	0	1

The nucleus of an atom consists of protons and neutrons. The protons repel each other with great electrostatic force. The neutrons, although of zero charge, help to keep the protons together in the nucleus. The force between protons and neutrons is called the strong nuclear force. Neutrons and protons are called nucleons.

- According to the shell model, electrons orbit the nucleus of an atom in 'shells' at specific distances from the nucleus. The shells are given quantum numbers $n = 1, 2, 3, 4, \ldots$, starting from the innermost orbit. The maximum number of electrons allowed in a shell numbered $n = 2n^2$. The most stable configuration of electrons in an atom occurs when the outermost shell is full. The shell model does not work well for atoms with many electrons.
- When atoms bond together, one type may gain and the other type lose electrons (electrovalent bonding); atoms may share electrons mutually (covalent bonding) or form an ionic lattice through which electrons meander (metallic bonding).
- In the Periodic Table, elements are arranged in an atomic number sequence. The horizontal rows of the table are called Periods. Atoms of elements in the same Period have their outermost electrons in the same numbered shell. There is a repetition of chemical properties of elements across successive Periods.
- The vertical columns of the Periodic Table are called Groups. Each Group consists of atoms with the same number of outermost electrons. Thus, elements within a Group have similar chemical properties.
- Three numbers may be used in describing the make-up of a nucleus. The atomic number or proton number (Z) is the number of protons present. The neutron number (N) is the number of neutrons present. The mass number or nucleon number (A) is the number of protons and neutrons in the nucleus. Thus, $A = Z + N$.
- Nuclides are species of nuclei with the same atomic number, Z. The symbol for a nuclide of an element, X, is written ${}^{A}_{Z}\mathrm{X}$.
- Elements whose nuclides have the same atomic number Z but different mass number A are called isotopes. Isotopes show identical chemical behaviour among themselves.

Answers to ITQs

ITQ1 $n = \frac{Q}{e} = \frac{1}{e} = \frac{1}{1.6 \times 10^{-19}} = 6.3 \times 10^{18}$ electrons

ITQ2 Unstable. The nucleus has a large number (85) of protons.

ITQ3 Maximum number $= 2n^2 = 2 \times 3^2 = 18$

ITQ4 A series of dark (absorption) lines were seen in sunlight. They corresponded to lines expected for a then unknown element, now called helium.

ITQ5 Mass number is 35. Symbol ${}^{35}_{17}\mathrm{CL}$

ITQ6 A and C: ${}^{5}_{3}\mathrm{X}$ and ${}^{6}_{3}\mathrm{X}$

Examination-style questions

1 U represents one of the isotopes of uranium. How many protons are present in one atom of this isotope?

A 92

B 146

C 238

D 330

2 What will be the mass number of a neutral atom of thorium which has 90 electrons and 145 neutrons?

A 90

B 145

C 235

D 290

3 Where in an atom are the protons and neutrons?

	Protons	Neutrons
A	nucleus	orbits
B	nucleus	nucleus
C	orbits	orbits
D	orbits	nucleus

4 How many electrons are present in a charge of –1.2 μC (micro-coulomb)? ($e = 1.6 \times 10^{-19}$ C)

5 Answer the following questions with reference to the Periodic Table (figure 33.3 and table 33.3).

(i) Which element has an electron configuration of:

(a) 2, 8, 5

(b) 2, 8, 8, 2

(ii) Name an element that has chemical properties similar to magnesium (Mg).

(iii) In which Period and Group is the element silicon found? Write the electron configuration of the silicon atom.

(iv) Draw a shell diagram to show an atom of nitrogen (N).

6 Discuss the following statements:

(i) A nucleus has protons (positively charged particles) very close to each other. Yet, the nucleus is quite stable and does not fly apart.

(ii) Atoms are normally neutral, electrically.

(iii) Lithium (Li) is very reactive but neon (Ne) is not.

7 Which two types of particles contribute mostly to the mass of an atom?

8 What is a condition, within an atom, that favours:

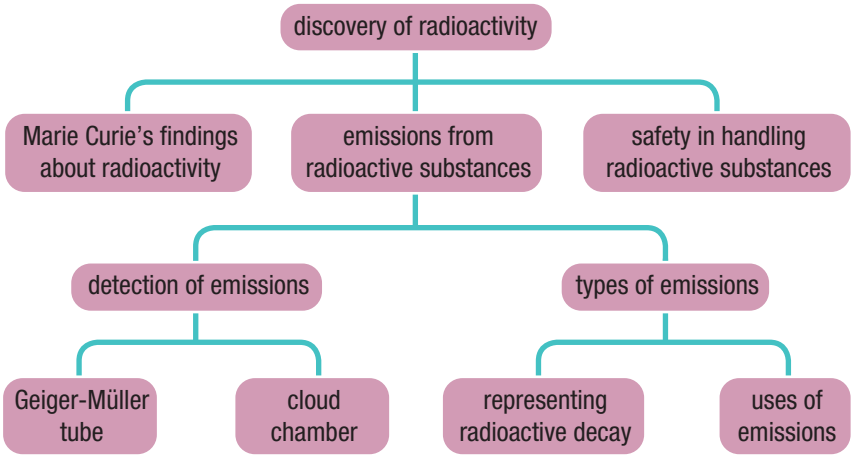
(i) a nucleus being most stable?

(ii) the electron arrangement being most stable?

34 Emissions from the Nucleus

By the end of this chapter you should be able to:

- describe Marie Curie's work in the field of radioactivity
- recall that the decay process is independent of conditions outside the nucleus
- discuss the need for safety measures in handling and using radio-isotopes
- describe how a cloud chamber works
- describe how a Geiger–Müller tube works
- recall the nature of the three types of emissions from radioactive substances
- describe experiments to compare the ranges of alpha, beta and gamma emissions in various media
- predict the effects of magnetic and electric fields on the motion of alpha and beta particles and gamma rays
- describe the appearance of the tracks of radioactive emissions in a cloud chamber
- associate radioactivity with nuclear instability
- represent and interpret nuclear reactions of the form
  ${}^{14}_{6}C \rightarrow {}^{14}_{7}N + {}^{0}_{-1}e$ (β particle)
  ${}^{226}_{88}Ra \rightarrow {}^{222}_{86}Rn + {}^{4}_{2}He$ (α particle) $(+ {}^{0}_{0}\gamma)$
- discuss the uses and dangers associated with emissions from the nucleus



The discovery of radioactivity

One cloudy day in 1896, Henri Becquerel, a French scientist, made an amazing discovery. He had rested some uranium samples on a covered photographic film in a closed drawer for four days. On developing the film, he noticed that patches of the film had been exposed as if light had struck it (figure 34.1). He guessed that the uranium sample must have been living off something invisible that had passed through the opaque cover of the film.

For a while, the emissions were called 'Becquerel's rays'. In 1899, experimenting in Canada, Ernest Rutherford showed that the emissions that

radiated from the uranium compounds were actually of three kinds. The emissions were given the names alpha (α), beta (β) and gamma (γ) radiation. The phenomenon of giving off these emissions became known as **radioactivity**.

radioactivity ›

The work of Marie and Pierre Curie

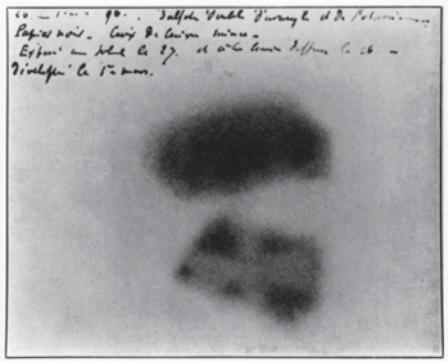


Figure 34.1 The appearance of a photographic film fogged by radiation from a uranium sample.

A few months after Becquerel's discovery, Marie Curie, a young chemistry student, and her physicist husband, Pierre (shown in figure 34.2), began investigating uranium compounds. The following are two of Marie Curie's major accomplishments:

1 She discovered that the amount of radiation emitted depended only on the amount of compound present. It did not depend on external factors such as temperature. In other words, radioactivity is determined solely by something within the atoms.

2 She discovered an element 300 times as radioactive as uranium. She named the element polonium, after Poland, where she was born. Marie and Pierre also discovered another element 900 times as radioactive as uranium. They named this element radium.

Marie and Pierre Curie along with Henri Becquerel were awarded the Nobel Prize in 1903 for their investigation of radioactivity. Marie was awarded a second Nobel Prize in 1911 for her discovery of polonium and radium. Marie was the first woman to win a Nobel Prize, and the only person to win two Nobel Prizes for work in science.

Marie Curie lived for 30 years after her first exposure to radiation from highly radioactive materials. Her continuous exposure to these materials resulted in burns to her hands. She suffered from continuous fatigue – one of the signs of radiation sickness. When she died, she showed many of the symptoms of **radiation sickness**. Others, experimenting with radioactive materials at the time of Marie Curie, did not live as long as she did following their exposure.



Figure 34.2 Marie Curie and her husband, Pierre, in their laboratory.

radiation sickness ›

Symptoms of radiation sickness

- Loss of body hair
- Nausea
- Fatigue
- Drop in white blood cell count
- Sterility after over-exposure

Caution!

Only persons trained to handle radioactive materials should be performing experiments with them. Others should keep well away from these materials. The international safety symbol warning of the presence of dangerous radiation is shown in figure 34.3.



Figure 34.3 Radiation warning sign.

Safety in handling radioactive materials

The experiences of Marie Curie and others showed that radioactive emissions cause considerable damage to living tissues. The facts that such emissions are invisible and that their effects are not immediately felt add to the deadly hazard they present.

Generally, the following precautions should be taken with radioactive materials:

1 Radioactive materials should be handled with tools, such as forceps – never with bare hands.

2 Sources should never be pointed towards people or other living organisms. Sources should never be brought near to the eyes.

3 There must be no eating or drinking where radioactive materials are being used.

4 After use, radioactive materials should be stored in lead containers and placed in a locked, isolated, metal cupboard.

5 After using radioactive materials, the user must wash his/her hands thoroughly before eating.

6 Equipment capable of detecting radioactivity must be present to check for radioactive emissions.

ITQ1

Can radioactive particles enter the body other than by touching or by eating substances that are radioactive?

Detecting radioactive emissions

radioactive emissions

Radioactive emissions can be detected by several methods:

1 The emissions fog photographic film. That is how they were discovered by Becquerel, in the first place.

2 When radioactive emissions strike certain crystals, they cause the crystals to give off flashes of light. It was these flashes, on a zinc sulphide screen, that Geiger and Marsden observed through their microscope (see page 499).

3 Radioactive emissions cause particles in their path to become ionised (electrically charged). When an emission collides with an atom, an electron may be knocked out of the atom. The remainder of the atom becomes a positive ion. Thus a negatively charged particle (the electron) and a positive ion are produced when the collision takes place. The detecting devices described below make use of the *ionising* effect produced by radiation.

The cloud chamber

cloud chamber

The low temperature within a **cloud chamber** enables a supersaturated alcohol vapour to form as alcohol evaporates. As vapour particles become ionised by collision with radioactive emissions, the ions formed (called condensation nuclei) attract other vapour molecules, which coalesce to form tracks of visible droplets.

Two designs of cloud chambers are shown in figure 34.4. The low temperature in figure 34.4 (a) is produced by the presence of solid carbon dioxide ('dry ice'). In the other design (figure 34.4 (b)), the low temperature is achieved during sudden expansion of the vapour when a piston, within a cylinder connected to the chamber, is pulled downwards.

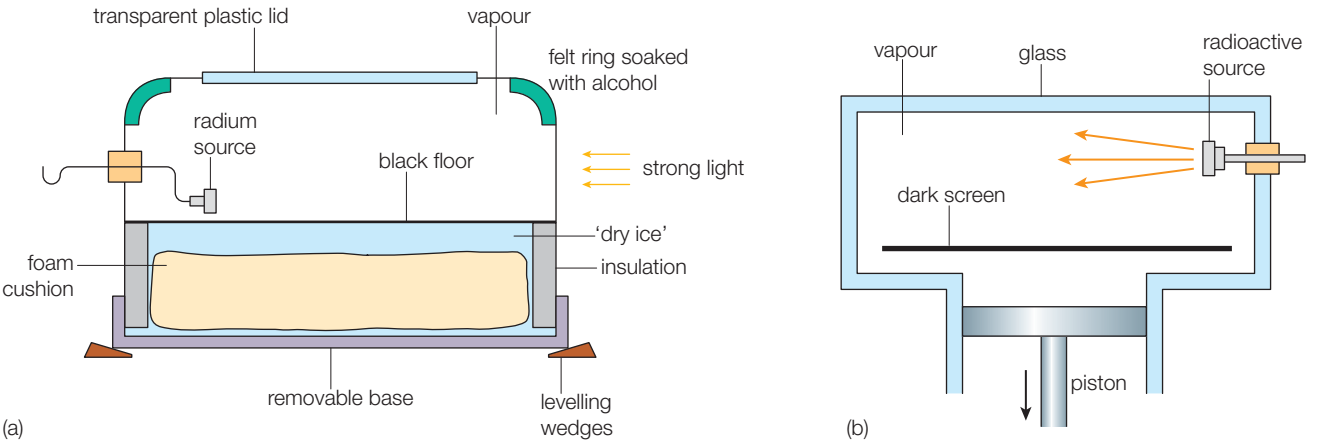


Figure 34.4 Two designs of cloud chambers.



Figure 34.5 Cloud chamber tracks of alpha particles are short, thick and grouped into different lengths.

Cloud chamber studies have shown that some emissions (alpha particles) give rise to short, wide tracks of almost equal length for a given nuclide (figure 34.5). Others (beta particles) result in longer, thin tracks, which are not necessarily straight. Yet others (gamma rays) give rise to discontinuous, thin, irregularly formed tracks. Further studies showed that the first two types of emissions consist of charged particles. The third consists of electromagnetic waves of high energy (photons). Tracks resulting from all three types of radiation are shown in figure 34.6.

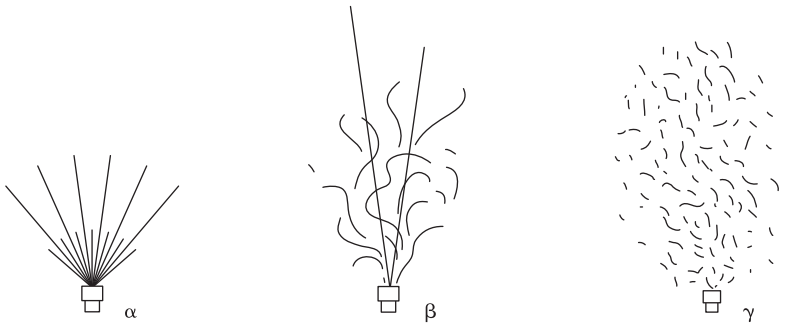


Figure 34.6 Cloud chamber tracks of α, β and γ radiation.

Cloud chamber tracks

Based on figure 34.6 above, discuss the following in your group:

1 How can we explain the observations that tracks of alpha particles are thick?
2 How do we explain why tracks of alpha particles of equal length?
3 How do we explain why tracks of alpha particles are very short?
4 Tracks of all alpha particles originate directly from the source; however, tracks of beta and gamma radiations do not. What is a possible explanation for this?
5 Why are tracks of beta particles so wavy and of different lengths?
5 Why are tracks of gamma rays so discontinuous and short?

The Geiger–Müller (G-M) tube

Geiger–Müller tube ›

The **Geiger–Müller (G-M) tube** contains an inert gas (e.g. argon or neon) at low pressure, along with a trace of bromine vapour (figure 34.7). The metal tube, made of copper, is earthed. A metal wire along its axis is connected to a high positive potential. A thin mica window at one end of the tube allows entry of radioactive emissions.

Counters and ratemeters

A counter, connected to a G-M tube, displays the number of particles detected by the G-M tube in a given time. A ratemeter displays the average count rate (counts per second).

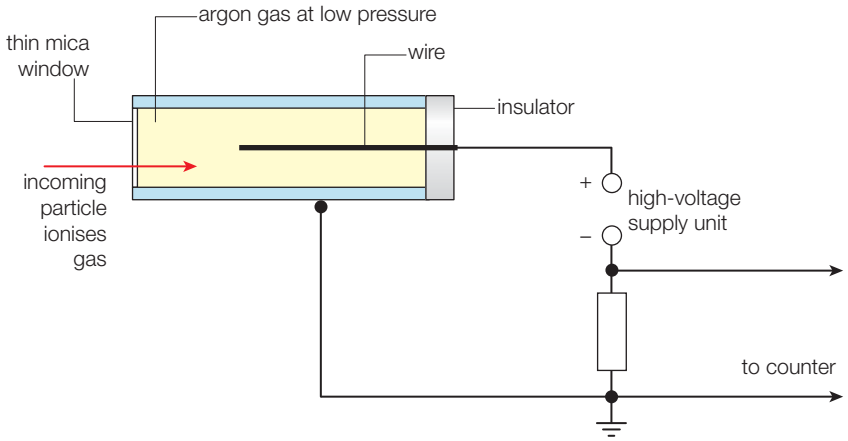


Figure 34.7 A Geiger–Müller tube.

When a charged particle or a photon passes into the tube, some of the argon atoms are ionised by collision with the in-coming radiation. The electrons released from the atoms rush towards the highly positive metal wire. On their way, the fast-moving electrons collide with other argon atoms, producing further ionisation. An avalanche of ion pairs thus results from a single radiation particle. When the argon ions reach the case, they are discharged since the case is connected to the negative of the power supply. A pulse of current flows around the circuit through the resistor, causing a voltage pulse across the resistor. Voltage pulses are sent to a **counter** (figure 34.8) or to a **ratemeter**.

ratemeter counter

The argon gas returns to its non-conducting state in less than a millisecond after the first collision. This quick return enables the G-M tube to detect individual particles coming at a fairly rapid rate.

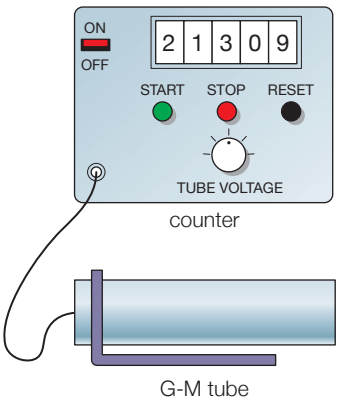


Figure 34.8 A G-M tube connected to a counter.

Random nature of radioactivity

A G-M tube connected to a counter (figure 34.8) can be used to demonstrate the random nature of radioactivity from a source. The count advances by 1 each time a particle of radiation or a gamma ray enters the tube. The counter numbers are observed to advance irregularly. This shows that the emissions from the radioactive source are given off randomly. (If a loudspeaker is present in the counting circuit, a click is heard each time the counter advances.)

Practical activity 34.1

Randomly occurring events

A random event is one which you cannot predict with certainty when it would happen. However, if the event is monitored a large number of times, a pattern often emerges, as this activity illustrates.

Method

1 Select 4 kings, 4 queens and 4 jacks from a pack of cards.



2 Decide on a particular "event" to be monitored throughout the activity (e.g. the pulling of the queen of spades).

3 Shuffle the 12 cards and have a partner pull one card without looking at the set. For each pull, record if the queen of spades was pulled.

4 Repeat steps 1 to 3 for 100 times.

Questions

1 What pattern did you observe in the results?

2 Predict the results for steps 1 to 3 carried out 200 times. Test your prediction.

Units of radioactivity

count rate

G-M tubes are often used for measuring count rates. The **count rate** is the number of ionising particles entering the tube per second.

becquerel (Bq) curie (Ci)

Two units of radioactivity are the **becquerel (Bq)** and the **curie (Ci)**, where

1 Bq = 1 emission per second

1 Ci = 3.7×10^{10} emissions per second

gray (Gy) ›

Biological damage and absorbed dosage of radiation

One gray (1 Gy) is about the same in energy as a sip of tepid coffee. Radiation sickness begins with a single dose of about 1 Gy. A single dose of 10 Gy usually causes death within days.

The biological damage done by absorbed radiation, however, does not only depend on the amount of energy absorbed (the number of grays). The damage also depends on the energy per particle, the kind of radiation, the number of particles entering a tissue per second as well as the type of tissue that is absorbing the radiation.

For medical purposes, another unit is used called the **gray (Gy)**. One gray (1 Gy) represents 1 J of energy absorbed from a radiation beam per kilogram of live tissue:

$$1\,\text{Gy} = 1\,\text{J kg}^{-1}$$

Radiation treatment of cancers usually involves carefully targeted absorbed doses of 40 Gy or more, delivered over a period of time. The aim of the treatment is to give as high a radiation dose as possible to the tumour and as low a dose as possible to the surrounding healthy tissue.

Personal dosimeters, worn by medical personnel, can detect and measure count rates in an area and/or total dosage of radiation that falls on parts of their bodies. Dosimeters come shaped as clip-on cards or small equipment.

For class discussion
What kinds of medical personnel should wear personal dosimeters?

Types of emissions from the nucleus

alpha particle | beta particle ›
gamma radiation ›

Further study of **alpha** (α), **beta** (β) and **gamma** (γ) radiation by Rutherford and others yielded information about the nature and properties of these emissions. This is summarised in table 34.1 (see also figures 34.6, and 34.10–34.12).

IT:Q2

1 Gy of radiation is absorbed from a beam of gamma radiation by the body. Which part of the body absorbs more of the energy – the flesh or the bones?

Table 34.1 The nature and properties of emissions from nuclei.

	Alpha (α) particles	Beta (β) particles	Gamma (γ) radiation
Nature	Positively charged particles, consisting of 2 protons and 2 neutrons. Thus, they are nuclei of helium atoms. Each has a charge of $+2e$ and a mass of approximately 4 u	These are electrons that come out of the nucleus. Each has a charge of $-e$ and a mass of approximately 0.0005 u	This consists of high-energy units of electromagnetic radiation (called photons). Photons have no mass or charge
Speed	They are ejected from nuclei at moderate speed. Most of the particles when coming from a single source travel with the same speed	The electrons are ejected at a variety of speeds: from near zero to close to the speed of light	γ rays travel at the speed of light
Range	They have a range of a few centimetres in air. They are also blocked by a thick sheet of paper and stopped by a thin sheet of aluminium foil	High-speed β particles may have a range of up to a few metres in air and several millimetres in aluminium	γ rays have no maximum range in air or any other material. About 1 cm of lead absorbs half of the γ radiation in a beam. So 2 cm of lead reduces the level of radiation in a beam to ½ × ½ = ¼. Note that no thickness of lead absorbs completely all the photons in a beam of γ rays

β emmission occurs when a neutron within the necleus undergoes a change to a proton (see page 524).

continued

	Alpha (α) particles	Beta (β) particles	Gamma (γ) radiation
Ionising power	They easily ionise air along their tracks by pulling electrons off air molecules. They lose their kinetic energy in the process and quickly come to a stop. In a cloud chamber, the tracks show up as short, thick lines of almost equal length	β particles have less ionising power than α particles. The tracks are fainter than those of α particles. Since they have small mass, most β particles are easily deflected by the outer electrons of atoms. So tracks tend to be irregularly shaped. Only very fast-moving particles might result in fairly straight tracks	No ions are formed along the path of a γ photon. In a given beam of photons, a single photon may interact with an atom, knocking out an electron in the process. The electron comes off with relatively slow speed and makes a short irregular track. Short tracks indicate where single photons have interacted with single atoms
Effect of electric and magnetic fields	α particles are deflected by electric and magnetic fields	β radiation is affected by electric and magnetic fields	γ rays are not affected by electric or magnetic fields

Experiment to compare the ranges of α, β and γ radiation

background count rate ›

The setup of the apparatus used for this experiment is shown in figure 34.9. First, a **background count rate** is measured, without the source, S, in place. (Average count rates are determined by dividing the total number of counts registered on the counter by the time for which the counts were measured.)

Internet search term: measuring radiation – YouTube

The general levels of background radiation are usually not high enough to cause harm.

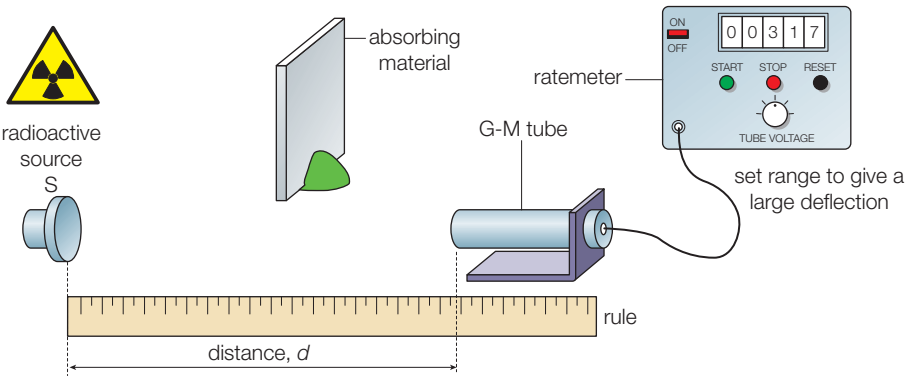


Figure 34.9 An arrangement that can be used to compare the ranges of α, β and γ radiation in various materials.

This background rate is due to radioactivity of substances present in the environment. It is subtracted from any other count rates measured so as to give a **'corrected' count rate**.

corrected count rate ›

To compare the ranges in air for α and β particles, each source of radiation is moved to different distances from the G-M tube and the corresponding count rates taken and then corrected. The distance corresponding to a corrected count rate of zero is the **range** of that radiation in air.

range ›

The ranges of α and β radiations in other materials are found by making the distance between the source and the G-M tube small and keeping it fixed. Different thicknesses of materials are placed between the source and G-M tube and corrected count rates taken. As before, the range in a material (figure 34.10) corresponds to the thickness of the material that just absorbs all the radiation, resulting in a corrected count rate of zero.

ITQ3

To what fraction of the original intensity will 4 cm of lead reduce γ radiation present in a beam? (Hint: see table 34.1 and figure 34.10.)

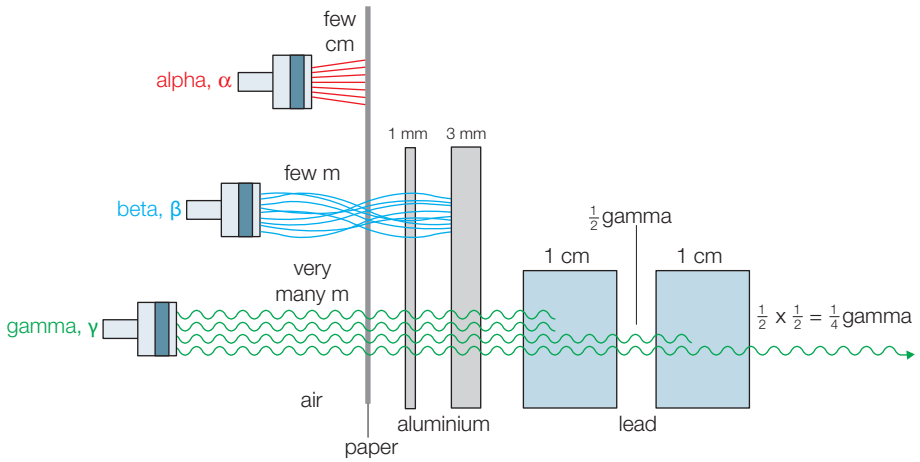


Figure 34.10 The ranges of radiations in different materials.

The range of γ radiation both in air and in other materials is indefinite. Hence, the thickness of material that will reduce the intensity of γ radiation in a beam by half is used instead to give an indication of the 'range' of fairly intense radiation in the material.

ITQ4

Why are alpha particles deflected much less than beta particles in the same field?

Composite diagrams

Figures 34.11 and 34.12 are composite diagrams showing results obtained separately with each type of radiation. Alpha particles are hardly deflected as compared to beta particles in the same field. The details do not show up as depicted – the diagrams are grossly exaggerated for illustration purposes.

Effect of electric and magnetic fields

Alpha particles are deflected away from a positive plate, as shown in figure 34.11. This shows that α particles are positively charged. Beta particles are repelled by a negative plate, showing that β particles are negatively charged. The fact that γ rays are not deflected shows that they have zero charge.

The effect of a magnetic field is shown in figure 34.12. The paths taken are in agreement with Fleming's left-hand motor rule (chapter 27) applied to charges moving at right angles to a magnetic field. The paths can be observed in a cloud chamber.

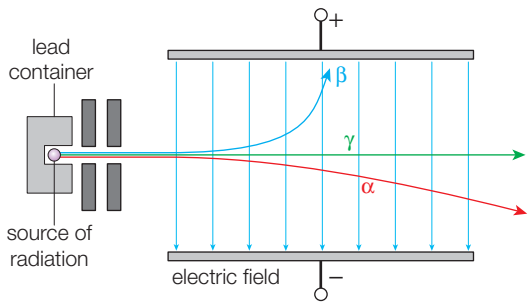


Figure 34.11 The effect of an electric field on the three types of radiation.

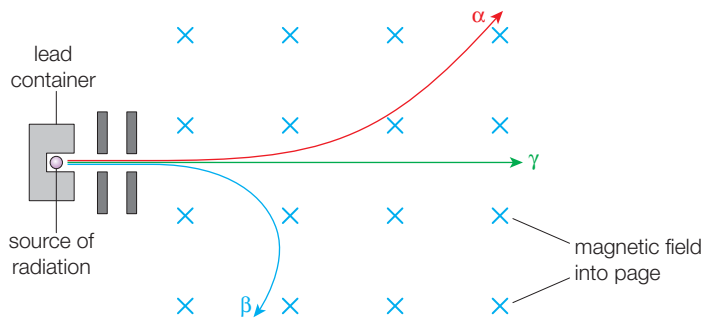


Figure 34.12 The effect of a magnetic field on the three types of radiation.

Representing radioactive decay

radioactive decay | unstable nuclei

When an atom emits α, β or γ radiation, we say that the atom undergoes **radioactive decay**. Radioactive decay occurs from **unstable nuclei** (refer to section on 'Stability of the nucleus' in chapter 33).

beta decay

Beta (β) decay

In chapter 33, we saw that for Z (the number of protons in a given nucleus) less than about 30, stable nuclei result when there are equal numbers of protons and neutrons present. Neutrons help to keep the protons together in the nucleus. However, nuclei with more neutrons than protons (in the same group $Z \leq 30$) tend to be unstable. Such nuclei tend to achieve stability by emission of a β particle from a neutron. The neutron becomes a proton in the process. Thus, the carbon-14 nucleus decays by β-particle emission (figure 34.13) and a more stable nitrogen nucleus is formed with equal numbers of protons and neutrons:

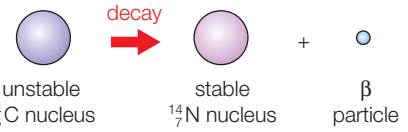


Figure 34.13 Illustrating β decay.

$$^{14}_{6}\text{C} \rightarrow {}^{14}_{7}\text{N} + {}^{0}_{-1}\text{e} \text{ (β particle)}$$

Beta decay also occurs in large nuclei where too many neutrons are present and contribute to instability.

alpha decay

Alpha (α) decay

Decay equations

Decay equations, such as the one shown in the main text, use the standard symbols for representing nuclides. The symbols are discussed in chapter 33, in the section 'Describing nuclides'.

For elements with $Z > 30$, more neutrons than protons are required for nuclear stability. This is because repulsive forces among protons can act over large distances within the nucleus. The forces provided by the neutrons, however, act over very short distances only. Hence more neutrons than protons are needed to hold the nucleus together when there is a large number of protons present within the nucleus. A nucleus that does not have enough of these neutrons will tend to be unstable. Such a nucleus usually achieves stability by emission of an α particle. Radium-226 decays to radon-222 by α-particle emission (figure 34.15). The number of protons in the nucleus is reduced, resulting in a somewhat more stable nucleus:

$$^{226}_{88}\text{Ra} \rightarrow {}^{222}_{86}\text{Rn} + {}^{4}_{2}\text{He} \text{ (α particle)}$$

gamma decay

Gamma (γ) decay

ITQ5

By how much does the mass number of a nuclide change after (i) alpha, (ii) beta and (iii) gamma emission?

Parents and daughters

A nuclide that decays is called a parent. The nuclide that results from decay is called the daughter. Thus, in the β decay equation given, carbon-14 is the parent and nitogen-14 is the daughter. The production of an element different from the parent is called **transmutation**.

For class discussion

What would the gamma decay equation for figure 34.14 be?

Beyond $Z = 82$, nuclei tend to be very unstable on account of the very large numbers of protons. The energy required to hold the protons together in such nuclei is large. Neutrons are no longer able to hold such nuclei in a state of stability. Apart from α-particle emission, γ rays may also be given off.

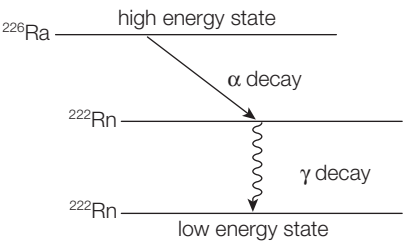


Figure 34.14 Energy level diagram showing γ emission followed by g emission.

Gamma emission usually occurs when a nucleus decays from an excited (high energy) state to a lower energy state. The decay of radium-226, shown in the previous equation, provides one such example. The radon-222 nucleus resulting from α decay is, at first, in a high energy state. The energy is lowered by emission of γ radiation. Figure 34.14 shows an energy level diagram depicting α emission followed by γ emission. Note that since γ radiation has no mass and no charge, the radon-222 nucleus does not change either in mass number or in atomic number following γ emission.

Making use of emissions from the nucleus

Generally, when making use of emissions from the nucleus, several factors are considered:

1 The first factor is safety. Nuclear radiation causes damage to living tissues. Alpha particles have a very short range. Thus, most of the damage they inflict is on surface layers of organisms. Beta and gamma radiation are more penetrating and can therefore result in internal injury.

2 The second factor relates to the length of time (or 'half-life;) the radiation keeps on being emitted. The 'half-life' of radioactive sources is discussed in chapter 38. In some applications, a long half-life is needed. In others, the long half-life of a radioactive source is considered a hazard.

3 The third factor relates to the desired function. Some devices depend on ionisation produced by radiation; others depend on heat produced; yet others depend on penetrating power; and finally some even depend on the ability to produce biological damage.

Radiation damage to living tissues

Alpha particles readily ionise living matter in their path. Thus, chemical reactions involving ions within cells are affected. Also, the energy given up as α particles come to a stop can cause tissues to overheat.

Beta particles also disrupt ionic reactions within living cells, though not as much as α particles do. However, β particles can cause damage to deeper-lying tissues.

Gamma radiation affects ionic reactions to a lesser extent than β particles. However, γ rays can penetrate deep into the body and can damage chromosomes, affecting cell functions, including reproduction.

Uses of nuclear emissions will be discussed in chapters 35 and 36 in terms of factors such as safety, half-life and desired function. The heart pacemaker and the smoke detector-alarm, both of which perform vital functions, and other applications discussed in those chapters, will show how useful nuclear emissions are, in spite of being so very dangerous.

Chapter summary

- Henri Becquerel, a French scientist, is credited with the discovery of radioactivity. Radioactivity is the giving off of emissions from the nucleus of an atom.
- Marie Curie's major accomplishments in studying radioactivity were:
  - the discovery that the amount of radiation emitted depended only on the amount of compound present;
  - the discovery of the highly radioactive elements of polonium and radium.
- Both the cloud chamber and the Geiger–Müller tube operate on the ionising effect of radiation. The cloud chamber shows tracks made by the radiation in cold air saturated with alcohol vapour. The Geiger–Müller tube is used for detecting individual particles or photons of radiation.
- Emissions from the nucleus are of three main types: alpha, beta and gamma.
- Alpha (α) particles are helium nuclei. An α particle has a charge of $+2e$ and an approximate mass of 4 u.
- Beta (β) particles are fast-moving electrons that come from the nucleus. As electrons, each has a charge of $-e$ and an approximate mass of 0.0005 u.
- Gamma (γ) photons consist of electromagnetic radiation. A photon has zero charge and a mass of 0 u.
- Both α and β radiation are affected by electric and magnetic fields; γ radiation is not. Both α and β radiation have definite ranges in matter; the range of γ photons is indefinite.
- Radioactive decay occurs from unstable nuclei. Instability may be caused by too many protons, or too many neutrons, in a nucleus.
- Alpha decay tends to occur from nuclei with a large number of protons.
- Beta decay usually occurs when the number of neutrons is large in relation to the number of protons in the nucleus.
- Gamma decay may occur when a nucleus is in a very high energy state. Gamma decay results in the nucleus achieving a lower, more stable energy state.

Answers to ITQs

ITQ1 Yes, by breathing in gases or particles in the air that are radioactive. Radon gas, for example, is radioactive.

ITQ2 Bone absorbs more energy from a beam of radiation than flesh because bone is denser than flesh.

ITQ3 Each cm thickness of lead reduces the level by half. Thus, 4 cm will reduce the level to $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$.

ITQ4 Alpha particles are deflected much less because an alpha particle has about 7000 times the mass of a beta particle but only twice the charge.

ITQ5 **(i)** decreases by 4; **(ii)** increases by 1; **(iii)** no change.

Examination-style questions

1 Which of the following lists nuclear radiations in order of increasing range in air?

A $\alpha\ \beta\ \gamma$ C $\gamma\ \alpha\ \beta$

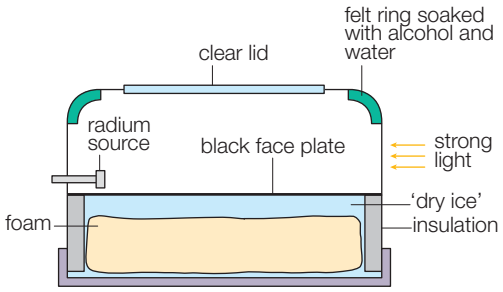
B $\beta\ \alpha\ \gamma$ D $\gamma\ \beta\ \alpha$

2 Which of the following gives the *best* description of α particle tracks in a cloud chamber?

A short, thick C long, thick

B short, thin D long, thin

3 The diagram shows a cloud chamber. A cloud chamber enables us to see the paths of emissions from radioactive sources.



(i) Describe the function of:

(a) the dry ice (solid CO_2),

(b) the alcohol and water,

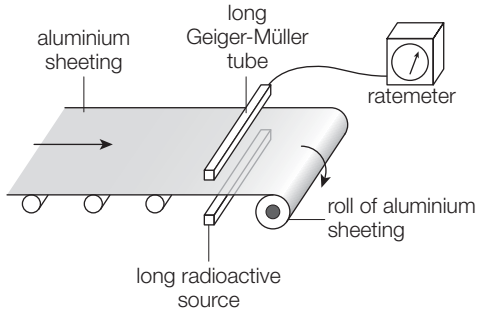
(c) the strong light and black face plate.

(ii) Draw diagrams to show tracks made by α, β and γ emissions.

4 (i) The table shows some of the properties of the emissions from radioactive substances. Identify A, B and C.

Type of radiation	Range in air	Penetration
A	greater than 1 m	passes through thin sheet of lead
B	6 cm	stopped by sheet of paper
C	70 cm	stopped by sheet of aluminium 3 mm thick

(ii) The diagram shows a method that is used to check on the thickness of aluminium sheeting during manufacture. The Geiger–Müller tube is used to detect radiation that has passed through the sheeting from a radioactive source.



(a) Give *one* reason why α particle radiation would *not* be suitable for the manufacturing process.

(b) Name a type of radiation that can be used and give *two* reasons for your choice.

(iii) Write an equation to illustrate:

(a) the decay of carbon-14 to nitrogen-14,

(b) the decay of ^{226}Ra to ^{222}Rn by alpha particle and gamma emission.

5 By how much does the atomic number of a nuclide change after

(i) α decay?

(ii) β decay?

(iii) γ decay?

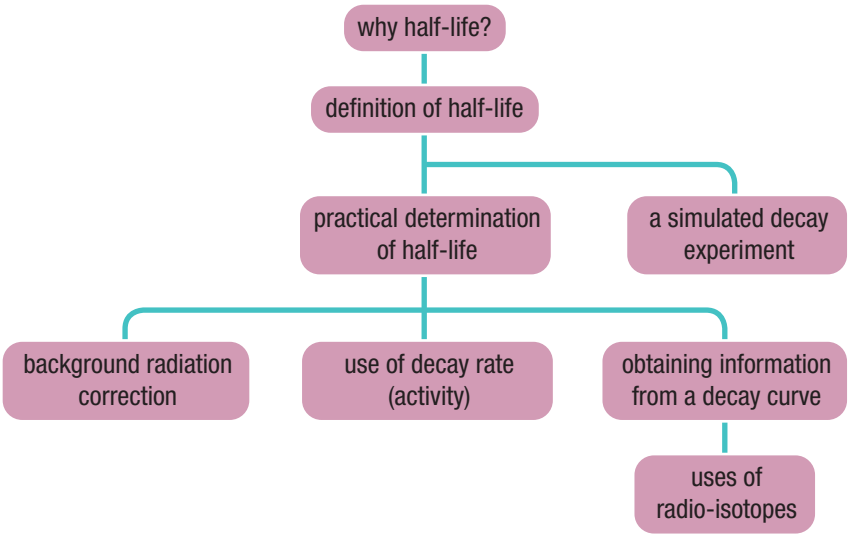
6 The table lists the properties of the three types of radiation, α, β and γ. Complete the missing properties in the table.

Properties	α particles	β particles	γ rays
charge	$+2e$		0
nature		electrons	
affected by magnetic field		yes	
approximate range in aluminium	0.1 mm		
degree of ionisation produced	very strong		

35 Half-Life

By the end of this chapter you should be able to:

- describe an experiment from which a radioactive decay curve can be obtained
- define the term 'half-life'
- perform analogue experiments to illustrate random 'decay' processes
- use graphs of random decay to show that such processes have a constant half-life
- work through simple problems on half-life
- discuss the useful applications of radio-isotopes



Why half-life?

radio-isotope

Radioactive isotopes of elements (called **radio-isotopes**) give off emissions and do so at different rates. Some take a long time before all activity practically ceases – others take a very short time.

If is often useful to know how long a radio-isotope sample would keep giving off emissions – in other words, the 'life' of the radio-isotope. Unfortunately, the 'life' of a decaying element cannot be precisely determined. This is because radioactivity is a random process. One nucleus within a sample might decay in 2 minutes' time. A neighbouring nucleus might decay 2 million years later! We can never be sure exactly when all nuclei within a sample have decayed. Since even a tiny sample may contain billions of atoms, the time

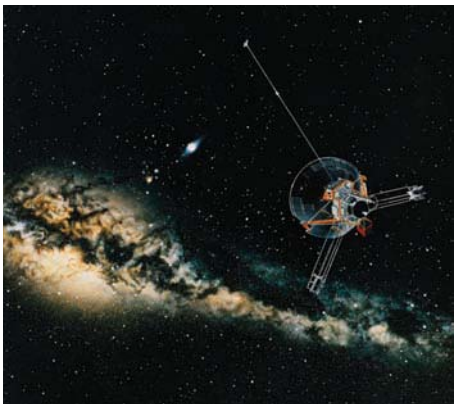


Figure 35.1 Pioneer 11 in deep space on its mission to explore the outer solar system (where solar power is very weak). Batteries powered by long "life" radioisotopes had to be used to run the electronics on board. The batteries were positioned on the two outstretched arms of the spacecraft (see page 546).

half-life ›

taken for *half* of the sample to decay is determined more easily. This decay time is called the **half-life**, and is the subject of this chapter.

Half-life

radioactive decay ›

Radioactive decay

We refer to an atom as having undergone decay if its nucleus has decayed, i.e. has given off some kind of radiation (see chapter 34, section 'Types of emissions from the nucleus', for kinds of radiation emitted).

Every radioactive isotope (or radio-isotope) has its own definite decay rate, which is characterised by its half-life:

> **The time taken for half the number of nuclei in a given sample of atoms to decay is called the half-life of the isotope.**

Half-lives vary from millionths of a second to billions of years! A radio-isotope with a long half-life decays very slowly. It would take a very long time for half the number of atomic nuclei present in a sample of such an isotope to undergo decay. Table 35.1 lists the half-lives of some selected radio-isotopes.

Table 35.1 Half-lives of selected radio-isotopes.

Isotope	Symbol	Half-life	Type of emission
polonium-214	Po	1.64×10^{-4} s	α, γ
radon-220	Rn	52 s	α
sodium-24	Na	15.0 h	β, γ
iodium-131	I	8.04 days	β, γ
cobalt-60	Co	5.27 years	β, γ
carbon-14	C	5.73×10^{3} years	β
americium-243	Am	7.37×10^{3} years	α, β
uranium-235	U	7.04×10^{8} years	α, β
thorium-232	Th	1.4×10^{10} years	α

ITQ 1

Which of the isotopes in table 35.1 would decay almost completely, i.e. would practically cease emission, within 1 second? Explain your answer.

Decay rate

Decay rate is measured in becquerel (Bq), where 1 Bq = 1 emission per second (see page 520). Decay rate is also called 'activity'.

Practical determination of half-life

Use of decay rate

decay rate activity ›

It is difficult to tell at a particular time how many atoms in a given sample have, or have not yet, decayed. Hence, **decay rate** (or **activity**) rather than numbers of atoms is used instead to determine half-life. We use decay rates as an estimate of the number of undecayed atoms. This assumption follows from Marie Curie's discovery that the amount of radioactivity is proportional to the amount of radioactive substance present, i.e. to the number of undecayed atoms present.

The rate of radioactive decay can be measured by a Geiger–Müller (G-M) tube attached to a counter or to a ratemeter (see page 519).

Background radiation correction

background radiation ›

Before attempting to measure the count rates of decaying atoms, the 'background' radiation rate must first be determined. **Background radiation** is always present (see page 522).

The average value of the background rate is taken over a time equal to the duration of the experiment. This *average* value is found, since radioactivity is a random process. The background radiation rate is measured at the site of the experiment and with the G-M tube pointing in the same direction as during the experiment.

ITQ2

A G-M tube, attached to a counter, detects 615 particles of radiation from a certain source in 15 s. What is the average activity (or count rate) in Bq?

Did you know?

Apart from materials present within the Earth, background radiation comes from materials present in our bodies, too! Also, the Sun gives off protons, which interact with particles in the atmosphere to produce showers of both particles and electromagnetic radiation.

ITQ3

Why is an α-particle source, rather than a β or a γ source, usually used in laboratory demonstrations of half-life?

ITQ4

How is radon-220 gas produced from thorium-232?

This average background radiation rate must be subtracted from radiation rates obtained from the counter or ratemeter to give the actual radiation rates being detected from a sample. Hence the term 'background radiation correction'.

Obtaining the decay curve data

The apparatus shown in figure 35.2 can be used to show how half-life can be measured. Radon-220, an alpha-particle emitter with a convenient half-life of 52 s, is used as the source. It is produced in the decay of the thorium compound in the generator bottle. Some radon-220 gas is squeezed from the bottle into a filter flask. The clips are closed. The reading on the ratemeter begins to rise.

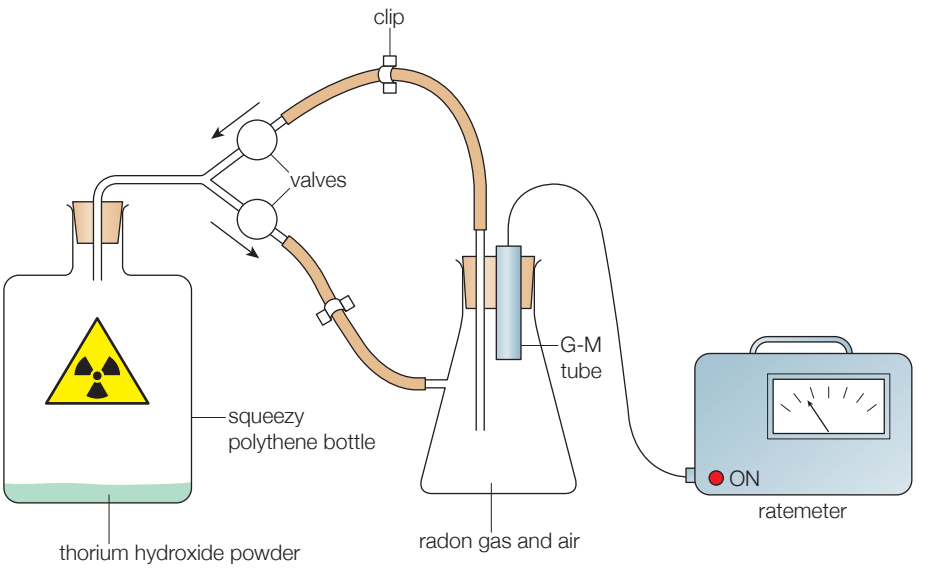
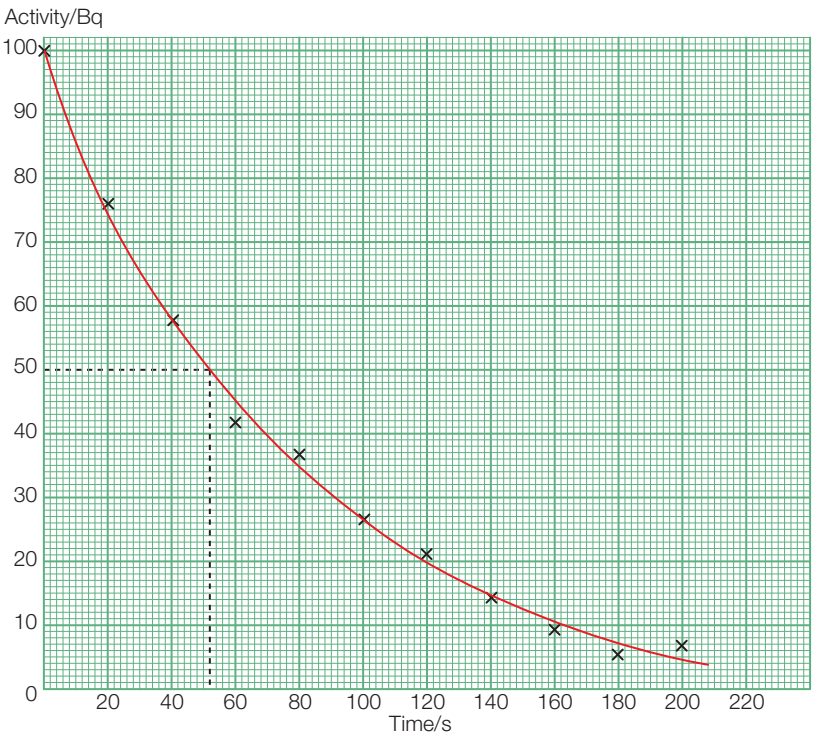


Figure 35.2 An arrangement for determining the half-life of radon-220.

When the ratemeter reading begins to fall, the count rate is taken every 20 s for about 3 minutes. The corrected rate (observed rate minus background rate) is plotted against time. The time taken for the corrected rate to drop to half of its original value is read from the graph. This time represents the half-life of the radon-220 sample used. A typical graph of corrected rate (activity) against time is shown in figure 35.3. Such a graph is called a decay curve.



decay curve

Figure 35.3 Graph of activity against time for a radon-220 sample. This is a **decay curve**.

At time $t = 0$ s, the activity is 100 Bq (1 Bq = 1 decay/s). The time taken for the activity to fall to half this value is the half-life of the radioactive element. From the graph, the fall from 100 Bq to 50 Bq takes 52 s. Thus (for this starting point), the half-life of the radon sample is 52 s.

Constancy of half life

The half-life of radioactive nuclide is a constant which is characteristic of that nuclide. Thus, the experimental value of half-life does not depend on the initial amount of radioactive substance or the initial value of the count rate chosen. Typical data obtained from a decay graph (Practical activity 35.1) illustrates this.

Practical activity 35.1

MATHEMATICS: graphical analysis

Half-life

Complete table 35.2 using data read from the decay graph in figure 35.3. Then answer the questions below.

Initial value of count rate, N/Bq	Half of initial value of count rate, $N_{\frac{1}{2}}$/Bq	Time taken for count rate to fall to half of initial value, $T_{\frac{1}{2}}$/s
100	50	52
80	40	54
70	35	
60		
50		
40		
mean $T_{\frac{1}{2}}$		

Table 35.2 Decay data.

Questions

1 What did you obtain as the mean value of the half-life, $T_{\frac{1}{2}}$?

2 Does the half-life depend on the starting mass (and hence, activity) of the sample used?

A simulation can also be done to show that the experimental value of 'half-life' does not depend on the initial amount of 'radioactive substance' or the initial value of the 'count rate' chosen. Practical activity 35.2 illustrates this.

Practical activity 35.2

A simulated 'decay' experiment

You will need:

- 100 equal-sized wooden cubes or dice (or, 100 1 cm lengths of pencil with hexagonal cross-section) each with one face blackened
- box of at least five times the volume of the cubes.

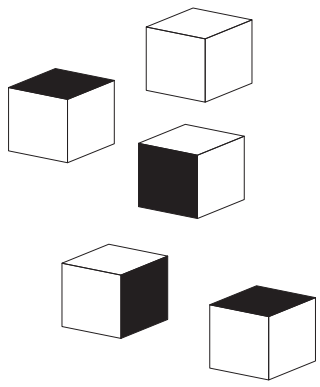


Figure 35.4 The cubes lying on a surface with the black face uppermost are treated as 'nuclei' of atoms that have 'decayed'. (Note: If you used dice, you could count the '6' as the black face.)

Method

1 Shake up the cubes well in a box.
2 Before spilling them on to a table, i.e. at spill 0, regard the number of decayed atoms as zero, since there are no cubes with a blackened face lying uppermost on the table. The number of undecayed atoms is therefore 100.
3 Spill the shaken-up cubes on to a table.
4 Remove the cubes with the black face uppermost from the table. (These represent nuclei that have 'decayed', see figure 35.4.)
5 Record how many cubes ('undecayed nuclei') remain on the table.
6 Put these remaining cubes in the box, and repeat steps 3–5, for at least 5 more spills.
7 Complete a table, showing number of spills from the box and number of 'undecayed nuclei'.
8 Plot a graph of 'number of cubes remaining' (N) against 'number of spills' (S).

Questions

1 The 'half-life' of the sample (cubes of wood) is the 'time' taken (number of spills required) for the number of cubes remaining ('undecayed nuclei') to drop by half. Using five values obtained from your graph, determine the mean 'half-life' of your set of cubes in terms of the number of spills. Compare your result with the values obtained by other students.
2 Predict the 'time' (number of spills required) for the number of 'undecayed atoms' (cubes) to drop from 100 to 25. Check your prediction experimentally.

Report

Write a lab report of the experiment.

Worked example 35.1: A numerical example involving half-life

The half-life of a certain radio-isotope is 12 h. How much of 400 g of the radio-isotope will remain after 36 h?

Solution

Since the half-life is 12 h

radioactive mass remaining after 12 h = 200 g

after next 12 h = 100 g

after next 12 h = 50 g

Hence, after 36 h (i.e. after 3 half-lives), 50 g of radioactive mass will remain.

ITQ5

The half-life of a certain radio-isotope is 2.0 min. What fraction of the radio-isotope remains after 8 min?

Uses of radio-isotopes

Many uses of radio-isotopes are related to their half-lives, relative ionising powers, ranges and/or energies of the emitted radiations.

Alpha-particle emitters

The smoke detector-alarm shown in figure 35.5 uses an α emitter. A very weak source of americium-241 is used. An α source is used because α particles have a very short range and hence will not be able to reach the outside of the detector. They would thus not be able to come into contact with people in the room where the detector is installed. The americium-241 source is also chosen because it has a long half-life (7370 years), which contributes to the reliability of the detector.

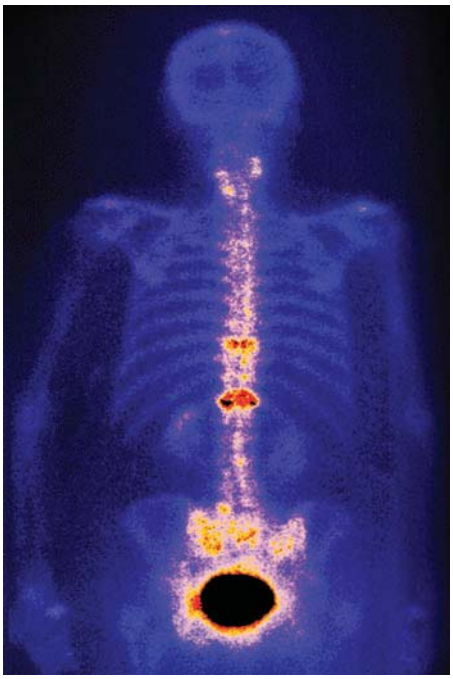


Figure 35.6 Iodine-131 radiograph showing absorption of iodine by the thyroid gland, in the neck, just below the larynx.

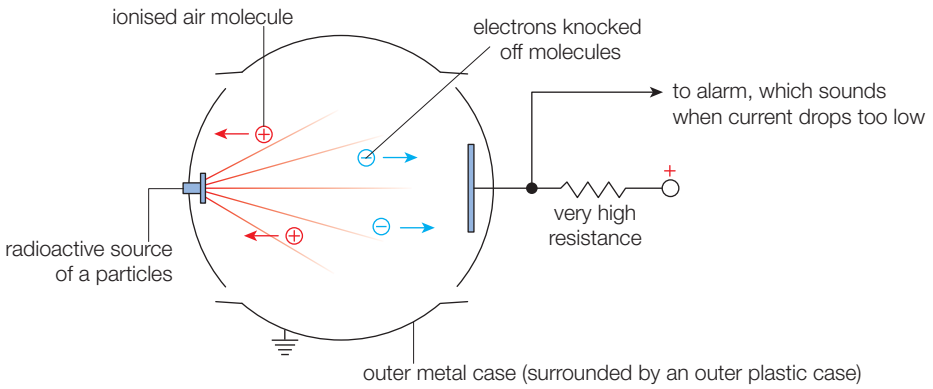


Figure 35.5 Inside a smoke detector-alarm.

The α particles ionise the air inside the detector and thus maintain a tiny ionisation current. As long as this current is maintained, the alarm does not sound. When smoke particles enter the detector, they absorb some of the α radiation. This causes the ionisation current to reduce and sets off an electronic alarm.

Some models of detector require a 9 V battery. When the battery becomes weak, the alarm chirps at regular intervals.

Batteries for distant exploration spacecraft (see figure 35.1) also make use of long half-life, α emitters. Inside the battery, the α particles generate heat when they come to a stop. The heat is converted to battery electricity, which operates the electronic circuitry in the spacecraft. Such a power source is called a radioisotope thermonuclear generator (RTG).

Beta-particle emitters

Small quantities of sodium-24 (a β emitter), in the form of sodium chloride, are injected into blood to serve as **radioactive tracers**. Beta particles can pass through living tissue but do not incur as much damage to living tissue as α particles do, since beta particles produce weak ionisation along their paths. A G-M tube, placed next to the skin, can be used to monitor the blood flow by detecting the β particles given off along the circulatory path. For safety reasons, sodium-24 is also chosen because it has a relatively short half-life (15.0 h). Other β-emitting radio-isotopes used in biological tracing include iodine-131 for tracing the uptake of iodine by the thyroid gland in the neck (figure 35.6) and phosphorus-32 for tracing phosphorus uptake by plants.

radioactive tracer ›

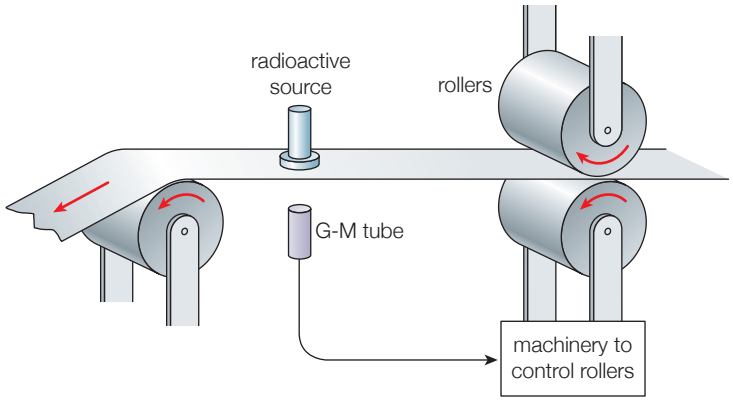


Figure 35.7 Thickness control of sheets using radioactivity.

Beta-emitting radio-isotopes of long half-life are used in thickness control during manufacture of paper sheets (figure 35.7). If the paper sheet is too thick, fewer β particles reach the detector. This sends a signal to a control device, which in turn sends a signal to the rollers to press the paper into sheets of the desired thickness.

Gamma-ray emitters

The high energy and penetrating power of γ rays have many uses. A very good source of γ rays is cobalt-60, which has a half-life of 5.27 years.

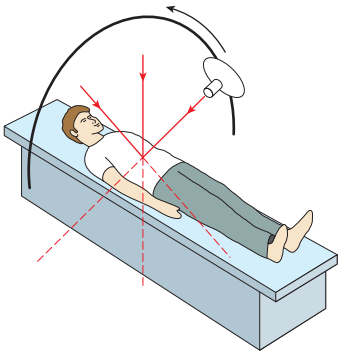
Gamma rays are used to 'see through' substances, including metals. Thus γ rays are used to inspect baggage at airports (figure 35.8). Gamma rays can also be used to detect hairline cracks or flaws in metals such as in an aircraft chassis. Gamma rays are also used for monitoring the thickness of metal sheeting during manufacture using a process similar to that shown for paper in figure 35.7.



Figure 35.8 Gamma rays are used in baggage inspection at airports.



(a)



(b)

Figure 35.9 (a) A γ radiotherapy machine used for destroying cancerous cells. (b) The source is rotated to minimise danger to healthy tissue and to direct the beam always towards the diseased tissue.

The Dead Sea Scrolls

Carbon-14 dating helped establish that the Dead Sea Scrolls (figure 35.10) were written around 100 BC (about 2100 years ago).

Radiotherapy makes use of γ rays, too. Gamma rays are aimed at cancerous tumours and kill cancerous cells (figure 35.9).

Gamma rays, since they can be lethal to living organism, are also used in sterilisation. An intense beam of gamma rays on foods including vegetables, meat, poultry and grains kills micro-organisms or else renders them sterile and thus enables long-shelf-life of the food. The need for chemical preservatives and refrigeration is thus greatly lessened. However, the rays do produce chemical changes in the food, mainly through ionisation of molecules.

Medical instruments are also similarly sterilised by gamma rays.

Group discussion
Uses of radio-isotopes
In your groups, discuss the following:

1. **Instead of using the long half-life radioisotope, americium-243, for deep space travel, couldn't solar voltaic cells be used as batteries?**
2. **Could cobalt-60 be used as a biological radioactive tracer instead of sodium-24? Give *two* reasons for your answer.**
3. **Give one reason for and one reason against using radon-220 instead of sodium-24 into the blood stream to monitor kidney function (via excretion of urine). Which reason carries more weight in the decision to be made?**

Other uses of radio-isotopes

Long half-life radio-isotopes are used in archaeological dating of rocks or remains of organisms. Living plants absorb carbon dioxide from the atmosphere. This carbon dioxide contains the isotopes carbon-12 and carbon-14. Carbon-14 (^{14}C) is radioactive and has a half-life of 5730 years. Carbon-12 (^{12}C) is stable. When a plant dies, it no longer takes in carbon dioxide from the atmosphere. This causes the ratio of ^{14}C to ^{12}C in the dead plant to decrease. A decay curve can be used to help to determine the activity per gram of the plant fossil. This enables the determination of the ^{14}C to ^{12}C ratio, and hence the fossil's age.

Carbon-14 dating assumes that the ^{14}C to ^{12}C ratio in the atmosphere has been constant over the centuries. This assumption is not necessarily correct. Hence, the decay curve is adjusted using activity from trees of known age. (The number of tree rings present can tell the age of the tree, and hence the time when the first set of wood was laid down in it.)

Figure 35.11 shows a ^{14}C radioactivity dating curve, assuming a half-life of 5730 years for the ratio-isotope. (Note the use of ages of trees in the calibration of the graph.) Carbon-14 dating is, however, not accurate beyond 60 000 years due to the very low levels of radioactivity associated with low level of ^{14}C present in samples being tested.

Archaeological dating of rocks and mineral deposits make use of other radio-isotopes. Uranium-238, with a half-life of 4.5×10^9 years, decays to stable lead-206. By measuring the activity per gram of a rock containing uranium-238 (alongside lead-206), the age of the rock



Figure 35.10

can be inferred. The age of the Earth has been estimated at 4×10^9 years, using this technique.

Moving parts in machinery sometimes incorporate iron-53, a radioactive element. The erosion of such machinery parts can be measured by the activity of the material deposited in the lubricating oil.

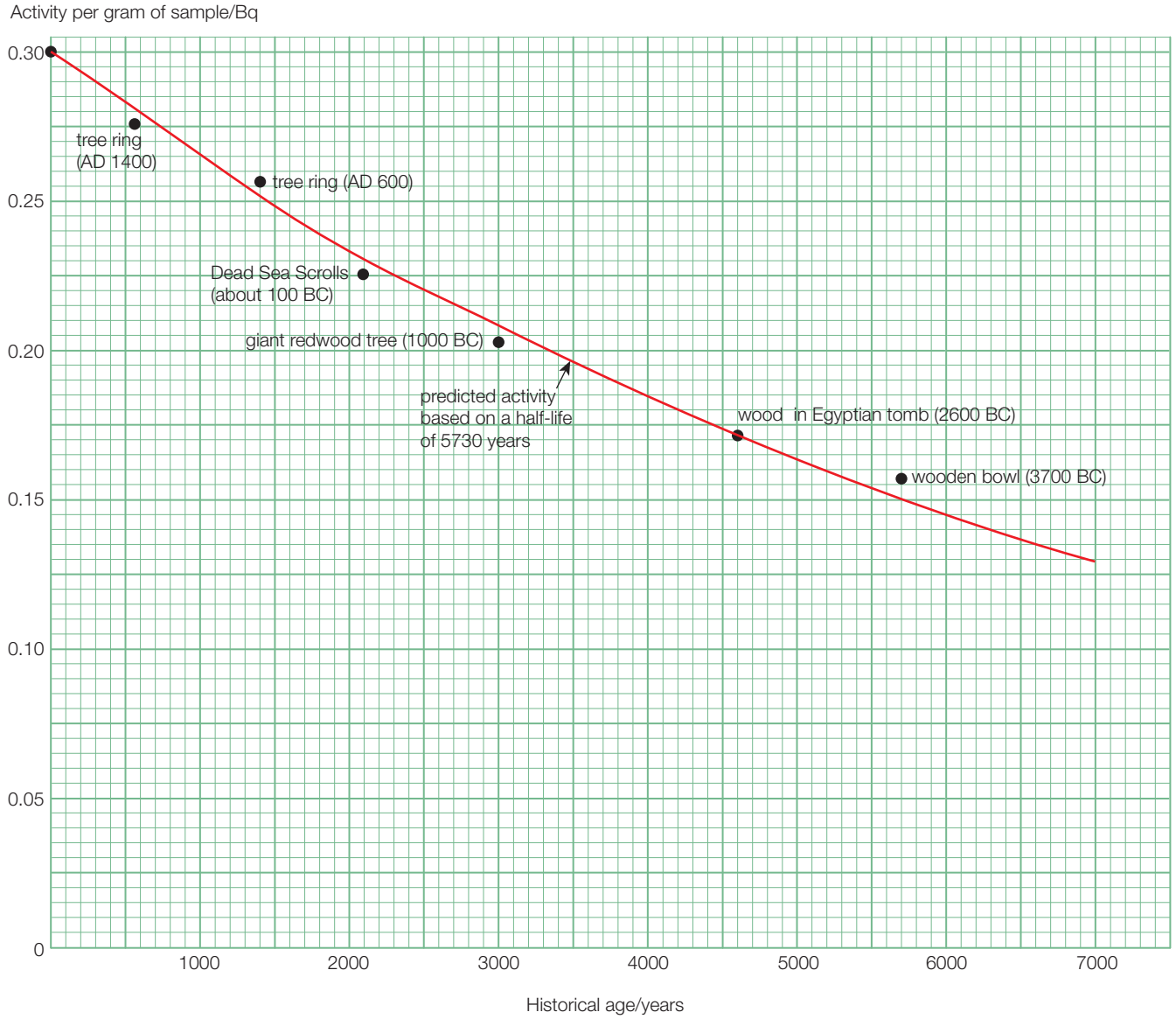


Figure 35.11 A carbon-14 radioactivity dating curve.

Chapter summary

- Because radioactivity is a random process, it is not possible to determine reliably how long it would take for a radioactive sample to decay completely. It is easier to determine how long it takes half of a given radioactive sample to decay.
- The half-life of a radioactive sample is defined as the time taken for half the number of nuclei within the sample to decay.
- Since it is difficult to determine the number of radioactive nuclei present in a sample, half-life is determined by measuring the activity (decays per second) of a sample. According to Marie Curie's discovery, activity is proportional to the amount of radioactive material present in a sample.

- The half-life of a given radio-isotope is constant for that isotope. Half-lives range from millionths of a second to billions of years.
- Alpha-emitting radio-isotopes are used in situations requiring short radiation range or large degree of ionisation.
- Beta-emitting (and sometimes gamma-emitting) radio-isotopes with relatively short half-lives are used as tracers in living organisms.
- Other uses of radio-isotopes include monitoring of thickness of materials during manufacture, archaeological dating, radiotherapy and sterilisation.

Answers to ITQs

ITQ1 Polonium-214. Since half of a ^{214}Po sample decays in approximately 10^{-4} s (0.0001 s), then nearly all of the sample will decay within 1 s.

ITQ2 Count rate in becquerels = number of particles detected per second

$$= \frac{615}{15} = 41 \text{ Bq}$$

ITQ3 Alpha particles have a very short range (a few millimetres) in air. So, they cannot travel far enough in air from the source to reach observers nearby.

ITQ4 The gas is one in a series of products of decay starting from thorium-232. The other products of this decay are solid, and thus the radon gas is easily separated from them. Thorium-232 has a long half-life (1.4×10^{10} year) and thus keeps on generating radon-220 gas through radioactive decay.

ITQ5 $\frac{1}{16}$. Fraction remaining after first 2 min = $\frac{1}{2}$

next 2 min = $\frac{1}{4}$

next 2 min = $\frac{1}{8}$

next 2 min = $\frac{1}{16}$

Hence, after 8 min, i.e. after 4 half-lives, 1/16th of the radioactive sample remains.

Examination-style questions

1 The activity of a radioactive sample is measured in which of these units?

A becquerel C watt

B joule D year

2 A corrected count rate of 100 Bq from a sample drops to 12 Bq in 15 h. The half-life of the sample is, approximately,

A 5.0 h C 10 h

B 7.5 h D 15 h

3 Refer to the radio-isotopes in table 35.1. Which of the following is the BEST isotope to use for monitoring the thickness of aluminium foil during manufacture?

A cobalt-60 C radon-220

B polonium-214 D sodium-24

4 According to the decay curve in figure 35.11, if 1 g of wood from an ancient bowl shows a ^{14}C activity of 0.12 Bq, what is the most probably age of the wood?

A 1600 years C 7000 years

B 5600 years D 7500 years

5 Which warning sign indicates the presence of radioactive materials?

A  B  C  D 

6 The table shows the half-lives of four isotopes of uranium. One gram of which of these isotopes would register the highest activity?

	Isotope	Half-life
A	uranium-232	68.9 year
B	uranium-233	1.6×10^5 year
C	uranium-238	4.5×10^9 year
D	uranium-239	23.5 min

7 (i) What is meant by the term 'half-life'?

(ii) Will a radioactive sample decay completely in two half-lives? Give a reason for your answer.

(iii) Radon-220 has a half-life of only 52 s. Yet radon-220 can be found naturally. Give an explanation for this occurrence.

8 The data in the table were obtained in an experiment to determine the half-life of a certain radio-isotope.

Background activity = 5 Bq							
Time/s	0	50	100	150	200	250	300
Observed activity/Bq	445	250	140	89	45	35	17
Corrected activity/Bq	440			84			

(i) Why was it necessary to obtain the background activity?

(ii) Copy and complete the table to show the corrected count rates.

(iii) Plot a graph of corrected count rate (*y*-axis) against time (*x*-axis). Use the graph to obtain *two* average values of the half-life of the radio-isotope.

(iv) Give *one* reason why the graph does not necessarily pass through all the points plotted.

9 A radioactive source and detector are used to control the manufacture of sheet steel of thickness 5.0 mm (in an arrangement similar to that in Figure 35.7). The following sources are available:

Radio-isotope	Radiation emitted	Half-life
radium-224	alpha	3.6 day
strontium-90	beta	28 year
barium-139	beta	85 min
cobalt-60	gamma	5.3 year
lead-211	gamma	36.1 min

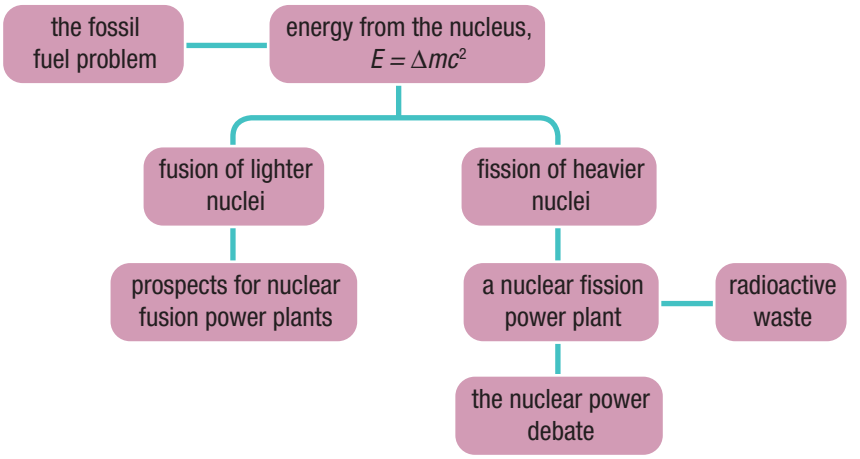
(i) Which of the sources listed above is *most* suitable for the manufacturing process described? Explain your choice of answer.

(ii) Name a suitable detector.

36 Energy from the Nucleus

By the end of this chapter you should be able to:

- associate the release of energy in a nuclear reaction with a change in mass
- work out problems using Einstein's equation $E = \Delta mc^2$
- relate the processes of fission and fusion to $E = \Delta mc^2$
- describe the main features of a nuclear power plant and the significance of critical mass
- discuss problems associated with the operation of a nuclear power plant
- discuss problems posed by radioactive waste
- advance arguments for and against the utilisation of nuclear energy



The energy problem of today

fossil fuels

Most of today's energy for home, transport and industry comes from **fossil fuels** (see figure 36.1). Gasoline, diesel, aviation fuel, LPG (cooking gas) and bunker C fuel (for power plants and heavy industry) are all refinements from fossil fuels, such as crude oil. Coal is used in power plants and in heavy industry. Fossil fuels take millions of years to form. However, there are two major concerns about the continued use of these fuels.

Fossil fuels
Fossil fuels are formed when the remains of plants and other organic material decay under pressure beneath layers of earth. Coal, crude oil and natural gas are fossil fuels. These fuels took millions of years to form.

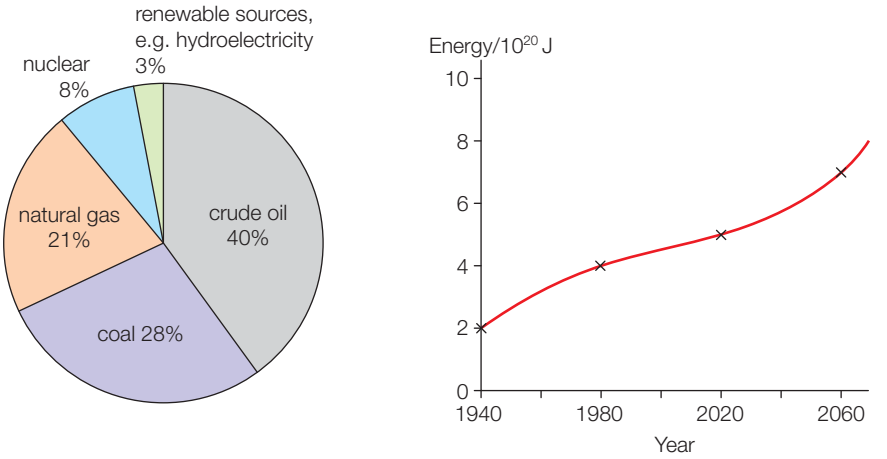


Figure 36.1 World dependence on fossil fuels.

pollution › acidic gases greenhouse gases ›

First, there is the chemical **pollution** that arises from the combustion of fossil fuels. **Acidic gases** such as sulphur dioxide, **'greenhouse' gases** such as carbon dioxide, and toxic gases such as carbon monoxide are three of the gases given off when fossil fuels undergo combustion to produce energy. Second, the fuels are being used up at a much faster rate than they are being formed. According to some predictions, based on current world population trends and the associated rates of 'development' anticipated by the populous developing world, and hence, the increasing needs for energy, the world's supply of economically available crude oil and natural gas is expected to last only until to about 2050. This means that it is likely that your children might face a serious fossil fuel energy shortage in their lifetime. Some authorities are optimistic, however. Every now and then, a new reservoir of oil is discovered. New technology and other fossil fuel sources, e.g. shale, might be developed which would allow the use of reserves that are at present uneconomic (but this does not solve the pollution problems arising from fossil-fuel usage). Already, too, there is serious opposition building up to a new technology to extract fossil fuel, called 'fracking', on account of environmental concerns. Economically available coal is expected to last for another thousand years, but so are the pollution problems that go along with its use.

Greenhouse gases

Greenhouse gases, such as carbon dioxide, are believed to be contributing significantly to the warming up of the Earth – a phenomenon called global warming. Severe changes in weather patterns are being predicted on account of this global warming trend.

BIOLOGY & ECOLOGY: environmental pollution

Internet search term: fracking

Class discussion

1 What are some factors that might be causing the very rapid rate at which oil and gas are being used up in this century?
2 When fossil fuels are being used, what dangers to the environment are posed by the following products: (a) acidic gases, (b) greenhouse gases, and (c) toxic gases?

ITQ1

According to figure 36.1, which form of fossil fuel is being used the most in the world?

Alternative sources of energy are therefore needed. Some being developed make use of ocean waves, water stored at heights, wind and even sunlight. But in these sources, the energy is very spread out. Vast areas of land or water space are required in order to collect the amounts of energy needed in the world today – posing technological, environmental, housing and space challenges for crop-growing.

A nuclear energy alternative?

nuclear energy ›

One very promising source of energy, already in use, is **nuclear energy**. When a relatively small amount of nuclear mass is converted into energy, a tremendous amount of energy is released. In nuclear fission, energy to satisfy the world's demands is estimated to be available for at least 1000 years (figure 36.2). But, as with the use of fossil fuels, there are environmental and public health dangers associated with this option. Nuclear fusion could provide the world's energy needs for a practically unlimited length of time, but has yet to be controlled. We shall examine these two possibilities of energy supply in this chapter.



Figure 36.2 The promise: nuclear energy, properly managed, can power large cities for a long time. This power station is Sizewell B, in the UK.

Energy from the nucleus

mass-energy ›

Einstein and mass-energy

ITQ2

On average, each person in the 'developed' world uses about 5000 J of energy every second. Assume that 20 MJ of energy is available per kilogram of fossil fuel burned. Estimate the mass of fossil fuel needed to maintain such a person's lifestyle for one year.

In 1905, Albert Einstein, a German scientist (who migrated to the United States in 1939), proposed his Special Theory of Relativity. Among other things, this theory stated that mass and energy were interchangeable. His famous equation of the equivalence was

$$E = \Delta mc^2$$

where Δm represents the mass converted into energy, E, and c is the speed of light in a vacuum ($c = 3 \times 10^8\,\text{m s}^{-1}$). The mass, Δm, 'disappears' and is converted into energy. This equation is often written as $E = mc^2$.

Missing nuclear mass: a clue to mass-energy

mass defect ›

mass defect of a nucleus = mass of the nucleus – total mass of its constituent protons and neutrons

Careful measurements have since supported Einstein's theory. It was found, for example, that the total mass of the separate protons and neutrons that make up the helium nucleus was greater than the mass of the same protons and neutrons when they are bound together in the nucleus. So how does one account for this difference in mass (called a **mass defect**)? According to Einstein's theory, this 'mass defect', Δm, represented energy within the helium nucleus. The value of this energy E is given by Einstein's mass-energy equation $E = \Delta mc^2$ (given above).

Worked example 36.1: How much energy from 1 kg converted mass?

If 1 kg of matter is changed entirely into energy, how much energy will be produced?

Solution

We have $E = \Delta mc^2 = 1 \times (3 \times 10^8)^2 = 9 \times 10^{16}\,\text{J}$

Worked example 36.2: Power for how many households?

Suppose the average Barbadian household uses 5000 kWh of energy per year (i.e. $5000\,\text{W} \times 1000 \times 3600\,\text{s} = 180 \times 10^8\,\text{J}$ per year). How much such households could be powered for one year by 1 kg of converted matter?

Solution

We have, using the value calculated in Worked example 36.1,

$$\text{number of households} = \frac{9 \times 101^6}{180 \times 10^8} = 5 \times 10^6$$

That is, about 5 million households can be powered for 1 year by just 1 kg of converted matter! (The total population of Barbados is just over ¼ million.)

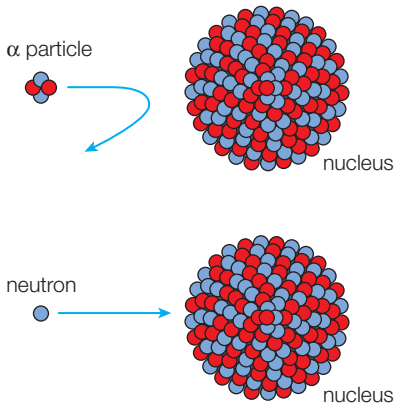


Figure 36.3 An α particle is repelled by a nucleus because both are positively charged. However, a neutron penetrates a nucleus easily since a neutron has zero charge.

Obtaining energy from the nucleus

Chadwick's discovery of the neutron (chapter 32) led to a series of important investigations with neutrons. Enrico Fermi (an Italian scientist who migrated to the United States in 1939), bombarded atoms of various elements with neutrons. He found that neutrons were easily 'captured' by nuclei (figure 36.3) since the former had no charge and were thus not repelled (as α particles were) from entering the nucleus. Fermi also made three important findings:

1 Isotopes of elements could be produced by bombarding the elements with neutrons. The isotopes are formed when the nuclei 'capture' neutrons. Some of the isotopes produced are unstable, and the nuclei formed split into fragments.

2 Slower neutron speeds tended to increase the changes of neutrons being 'captured' by nuclei.

3 Water placed in the path of moving neutrons slowed them down.

Nuclear fission

neutron capture ›

One example of a ^{235}U nuclear reaction involving **neutron 'capture'**, followed by splitting into fragments, is summarised below:

$$^{235}_{92}U + {}^{1}_{0}n \rightarrow {}^{236}_{92}U \rightarrow {}^{144}_{56}Ba + {}^{90}_{36}Kr + {}^{21}_{0}n + \text{energy}$$

Notice how the ^{235}U nucleus has 'captured' a neutron to become the unstable uranium-236 (^{236}U). The ^{236}U nucleus then breaks into two large pieces, two neutrons are released and tremendous energy is given off (figure 36.6 (a)). This breaking of a large nucleus into two smaller ones is called **nuclear fission**. Usually, highly radioactive products are formed during fission.

nuclear fission ›

Where does the energy come from? Careful measurements show that the total nuclear mass of the reactants ($^{235}_{92}U + {}^{1}_{0}n$) is greater than the total nuclear mass of the products formed ($^{144}_{56}Ba + {}^{90}_{36}Kr + {}^{21}_{0}n$). The apparent 'loss' in mass shows up as energy! Each of the two neutrons produced in the above reaction is also capable of initiating another reaction. Thus, a **chain reaction** (figure 36.4) becomes possible. A tremendous amount of energy could be obtained in this kind of reaction if many fissionable nuclei are present. If the reaction occurs in a very brief moment, then a very powerful *bomb* could result (figure 36.5). If the energy could be released slowly, it could be used to generate electricity and help alleviate the coming fossil fuel shortage.

chain reaction ›

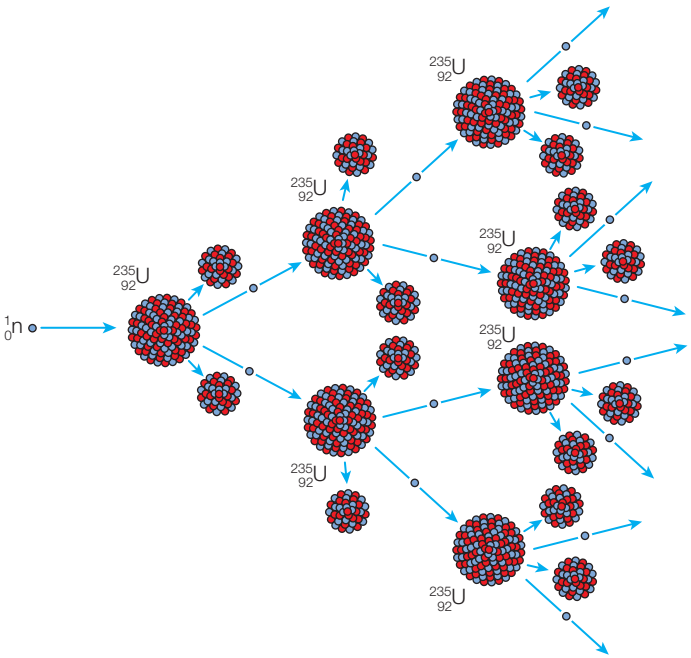


Figure 36.4 A chain reaction involving uranium-235.



Figure 36.5 The devastation of nuclear energy uncontrolled. An explosion such as this, caused by a nuclear bomb, destroyed the city of Hiroshima, Japan, and many of its inhabitants in an instant in 1945.

Nuclear fusion

nuclear fusion ›

Nuclear energy can also be obtained by **nuclear fusion**. When two very fast-moving deuterium (heavy hydrogen) nuclei are made to collide, fusion occurs to produce a helium nucleus (figure 36.6 (b)). Some mass 'disappears' and energy is released. Very little radioactive products are formed. Energy produced by fusion involves the use of the *lighter* elements such as hydrogen as opposed to heavier elements such as uranium.

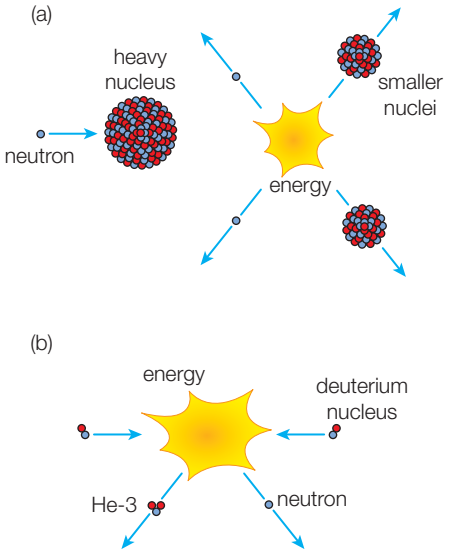


Figure 36.6 (a) Fission of a heavy nucleus caused by neutron capture. Energy is released in the reaction. (b) Fusion of two deuterium (heavy hydrogen, ^{2}H) nuclei to form a helium-3 nucleus (^{3}He). A neutron is given off and energy is released in the reaction.

Deuterium

The nucleus of an atom of ordinary hydrogen consists of one proton; the nucleus of 'heavy hydrogen' (deuterium) consists of one proton and one neutron. Deuterium is therefore an isotope of hydrogen.

Worked example 36.3: Energy given off in fusion

Use the masses in table 36.1 to calculate the energy released during the fusion of two deuterium nuclei described by the equation

$$^2_1\text{H} + {}^2_1\text{H} \rightarrow {}^2_1\text{He} + {}^1_0\text{n} + \text{energy}$$

Table 36.1 Mass data.

	Mass
deuterium (^{2}H)	3.3446×10^{-27} kg
helium-3 (^{3}He)	5.0084×10^{-27} kg
neutron (^{1}n)	1.6750×10^{-27} kg

Solution

From the equation for the reaction, we have

$$\text{mass of left-hand side} = 2 \times (3.3446 \times 10^{-27})\ \text{kg}$$
$$= 6.6892 \times 10^{-27}\ \text{kg}$$
$$\text{mass of right-hand side} = (5.0084 + 1.6750) \times 10^{-27}\ \text{kg}$$
$$= 6.6834 \times 10^{-27}\ \text{kg}$$

From these we get

$$\text{mass } \Delta m \text{ 'lost'} = (6.6892 - 6.6834) \times 10^{-27}\ \text{kg}$$
$$= (0.0058) \times 10^{-27}\ \text{kg}$$

Then using Einstein's equation gives

$$\text{energy } E = \Delta m \times c^2$$
$$= (0.0058 \times 10^{-27}) \times (3 \times 10^8)^2$$
$$= 5.22 \times 10^{-13}\ \text{J}$$

In Worked example 36.3, although the energy released per helium nucleus produced appears small, even small quantities of deuterium will contain billions of atoms and hence tremendous energy will result. One gram of deuterium could power the average modern Caribbean home for over 40 years!

The energy from the Sun is believed to originate from fusion of light elements like hydrogen. One estimate puts it that over 4×10^9 kg of mass is converted into energy each second in the Sun. However, the Sun's mass is so huge (about 2×10^{30} kg) that the Sun is not expected to 'die' for billions of years!

A nuclear fission power plant

An understanding of the structure and functioning of a nuclear fission power plant is useful in assessing the pros and cons of the utilisation of nuclear energy. In one design (figure 36.7), the power plant consists of:

nuclear reactor ›
- **nuclear reactor**;

heat exchanger ›
- **heat exchanger**;

steam turbine ›
- **steam turbine**;

generator ›
- **generator**.

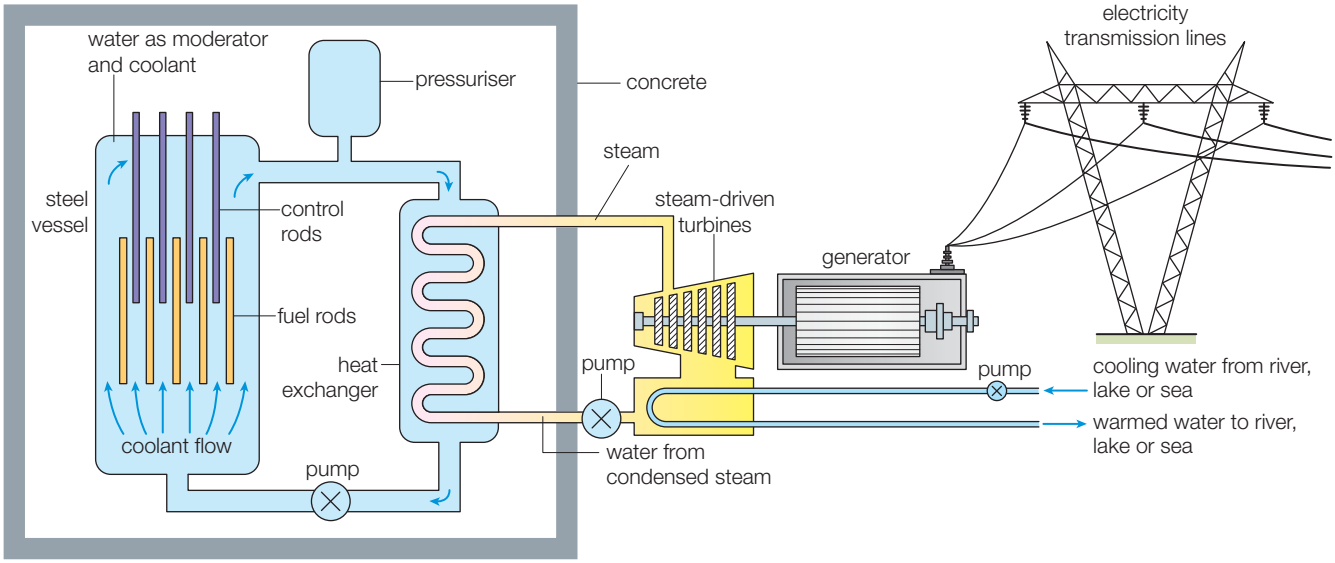


Figure 36.7 A power plant using a water-cooled nuclear fission reactor.

The nuclear reactor

Fuel rods

fuel rod ›

Fuel rods are rods containing uranium oxide. The uranium-235 in the oxide is the fuel of the reactor. Enriched uranium containing 2.3% ^{235}U, artificially increased from 0.7% in the natural state, is used in some reactors. Uranium enrichment is done by using high-speed centrifuges which separate uranium-235 from other non-fissionable isotopes of uranium.

A low percentage of ^{235}U used reduces the probability of neutron capture by ^{235}U, which is the fissionable material. This, in turn, reduces the likelihood of an uncontrolled chain reaction (figure 36.4). However, a certain minimum mass of ^{235}U, called the **critical mass**, is needed for a self-sustaining chain reaction to proceed.

critical mass ›

Many rods of the fuel are used rather than one solid block to be able further to control the reaction.

Moderator

moderator ›

The **moderator** surrounds the fuel rods. The moderator serves to slow down neutrons escaping from fuel rods so that they can be 'captured' by the uranium-235 nuclei in other fuel rods. (Neutrons come off with high speed when the ^{235}U nuclei split. For fission to take place, the neutrons must reach the ^{235}U nuclei at slow speeds.)

heavy water ›

In one design, heavy water serves as moderator. (A molecule of '**heavy water**' consists of two atoms of deuterium and one atom of oxygen. The deuterium atom is a hydrogen atom that has one proton and one neutron in its nucleus.) The heavy water also serves as coolant (see below). The advantage of a liquid moderator is that the liquid can be quickly drained out in an emergency. Neutrons would no longer be slowed down sufficiently for capture. Hence the chain reaction would slow down tremendously.

Control rods

control rod ›

The **control rods**, made of cadmium or boron, absorb neutrons. Thus, they can be raised or lowered to control the number of neutrons reaching the fuel and hence control the reaction.

Coolant

coolant ›

Either water or carbon dioxide gas pumped through the reactor serves as the **coolant**. The coolant carries heat energy generated at the core, the site of the fuel rods, to the heat exchanger. In the case of gas-cooled reactors, the pressure inside the reactor can reach as high as 40 atmospheres.

Steel pressure vessel

pressure vessel ›

The **pressure vessel** withstands the pressures built up in the reactor due to heated pressurised water (or heated gas). The concrete serves as a shield from radiation emitted from some of the radioactive by-products of the nuclear reaction. The concrete also shields from escaping neutrons.

Heat exchanger

The coolant gas (or liquid) carries the heat energy gained from the reactor core to a heat exchanger. The pipes of the heat exchanger contain water, which is converted into steam at high pressure by the heat.

Steam turbine and generator

Steam leaves the tubes of the heat exchanger at high temperature and pressure. The steam forces the turbine blades to spin. The spinning turbine causes the rotor of the generator to spin, thus producing electricity.

ITQ3

Why are the Pickering (on Lake Ontario, Canada) and Three Mile Island (in the USA) nuclear power plants sited next to large bodies of water?

Accidents at fusion reactors

- *Three Mile Island, USA* (1979): The power plant reactor overheated due to a malfunctioning water pump and safety valve and nearly exploded. Gases and water that were radioactive had to be deliberately dumped into the environment to avert a much worse catastrophe.
- *Chernobyl, Russia* (1986): An explosion at the power plant broke the 1000 tonne roof and sent deadly radioactive gases into the air. The gases spread throughout Europe and beyond, causing death and radiation sickness.
- *Fukushima, Japan* (2011): A tsunami and earthquake damaged seawater pumps. Three reactors experienced a meltdown due to the lack of flowing

coolant. Explosions occurred resulting in radioactive material being discharged into the atmosphere, notably iodine and caesium. In 2013, high levels of a toxic radioactive isotope, strontium-90, were reported to have been found in groundwater at Japan's Fukushima nuclear plant.

Prospects for fusion reactors

Energy has been obtained by fusion from hydrogen. Both Russia and the United States have already tested the hydrogen bomb. However, it has not yet been possible to control the rate of nuclear fusion. Work is still proceeding in trying to build fusion reactors, especially since the amount of radioactive waste produced in fusion, as compared to fission, is minimal and the supply of hydrogen is abundant.

ITQ4

Why is it so difficult to build fusion reactors?

The nuclear power debate

Already, nuclear power accounts for about 8% of the world's power needs, and is increasing. The energy is obtained by fission. A question that arises, especially in view of the imminent oil and gas fossil fuel shortage, is this: should countries in the Caribbean be encouraged to resort to nuclear power at this time?

We shall look at some arguments from both sides of this question. You should use the knowledge you have gained in this chapter to judge which side is better.

Points in favour of the use of nuclear power

1 A large amount of nuclear energy is available. Supplies of uranium-235 are expected to last for at least 100 years. Use of nuclear power means less dependence on fossil fuels. The precious fossil fuel resources can instead by used for making substances such as plastics and chemicals rather than simply being burned away.

breeder reactor ›

2 Some designs of nuclear reactors (**'breeder' reactors**) enable plutonium-239, another possible nuclear fuel, to be generated. Nuclear fuel could be available for several hundred years more by this process.

energy equivalent ›

3 Nuclear energy is a highly concentrated form of energy: 1 kg of ^{235}U fuel has the **energy equivalent** of 50 000 kg of fossil fuel. Thus, there are less transport costs. Fewer trips for fuel supplies are also needed. Land space required for an adequate supply of nuclear energy is minimal.

4 Nuclear energy generation, unlike energy generation from coal, does not result in the production of 'greenhouse' gases (such as carbon dioxide and methane), toxic gases (such as carbon monoxide) or acidic gases (such as sulphur dioxide).

5 Nuclear reactors are, at present, providing more than 8% of the energy being used in the world. Reactors have been operating in countries like the United States, Canada, the United Kingdom, France, Japan and Russia. Commercial reactors have been around since the 1950s.

Did you know?

A treaty signed in 1963 by 100 nations banned nuclear weapons testing in the atmosphere. This was done because it was recognised that radioactive gases are carried across the globe by air currents.

Points against the use of nuclear power

1 The effects of nuclear radiation are deadly. The radiation damages living cells and can lead to cancers and other abnormal growths. Humans can tolerate without series effects only a certain small dosage of this radiation per year. Nuclear radiation is invisible and is sometimes not even felt. This adds an insidious feature to its already hazardous nature.

2 Nuclear radiation is present at every stage in the process of obtaining energy: mining, transporting, enriching, milling the ore, refining the ore, and operating the reactor. Materials such as containers, clothing, buildings and equipment, exposed to radioactivity, become radioactive.

spent fuel ›

3 Waste disposal is a problem since the **spent fuel** (used fuel) is also highly radioactive. Apart from emitting deadly radiation, this waste still generates a lot of heat while disintegrating. One suggestion has been to encase the spent fuel in water in strong containers and bury them below the Earth's surface. But, according to some reports, storage tanks at a power station in Kent (UK) and in Harford, Washington (USA) have been leaking radioactive waste for some time. Hence, ground water could become polluted.

meltdown ›

4 The reaction in the reactor cannot be stopped abruptly. Thus there is the possibility of a **meltdown** when the reactor overheats due to failure of some part of the system. The overheating can cause an explosion, resulting in release of radioactive substances into the environment. Nuclear accidents involving excessive radiation leakage have occurred at Chernobyl, Russia, in 1986, Three Mile Island, USA, in 1979, at Windscale, UK, in 1957 and at Fukushima, Japan, in 2011. Radioactive gases sent into the air are carried about by air currents, carrying their dangers with them.

decommissioning ›

5 After about 50 years, a nuclear power station has to be decommissioned (closed down permanently). This is because parts of the plant would have become weak or radioactive. **Decommissioning** is a very costly and dangerous process, as it might involve removing highly radioactive material and pouring concrete over certain parts of the plant. Figure 36.8 shows Japanese protesters demonstrating for the safe decommissioning of a nuclear reactor.

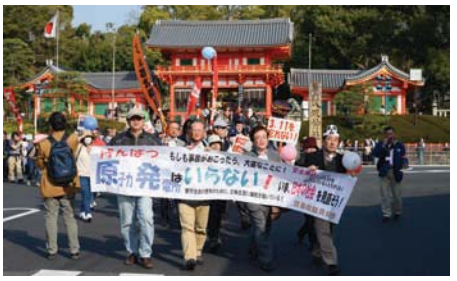


Figure 36.8 Japanese protestors demonstrating for the safe decommissioning of a nuclear reactor.

6 As for nuclear fuel (plutonium-239) being generated as a by-product of a nuclear reaction, the radiation dangers of plutonium are worse than those involved in using uranium-235. Plutonium-239 has a half-life of 24 000 years, and hence dangerous levels of radioactivity would be around for a far longer time! This will be so since only a fraction of fissionable fuel gets used up in a reactor. The remainder, which will contain large amounts of the fissionable material (plutonium-239 in this case), will present a serious disposal problem.

7 Using nuclear reactors might lead countries, and even individuals, or groups of individuals, to experiment with nuclear bomb production. There is also an increased chance of nuclear technology or materials being stolen. This nuclear proliferation could lead to the demise of our beautiful planet.

BIOLOGY & ECOLOGY:
environmental pollution

The challenge!

The Sun (figure 36.9) reminds us daily that nuclear energy from fusion is possible. How soon will we be able to duplicate this process in a controlled way on Earth and thus obtain an abundance of energy with little hazardous waste?

Internet search term: Iter fusion reactor in France

The promise revisited

Several of the big industrial nations, including the European Union, China, India, Japan, Russia, South Korea and the United States, are pooling their resources feverishly to duplicate nuclear fusion power in a controlled way on Earth. The project bears the acronym 'Iter' and the fusion reactor is being assembled in southern France. Initial costs are estimated to be £13bn or 15bn euros (US$21bn). If successful, an almost limitless supply of energy will be available to support activities on Earth and with practically little pollution. The source of the energy is likely to be a form of hydrogen, which is in vast supply in the water of oceans that cover over half of the Earth. Some optimistic estimates project that this type of power might be possible commercially by about the year 2075!

And now … decide your position.

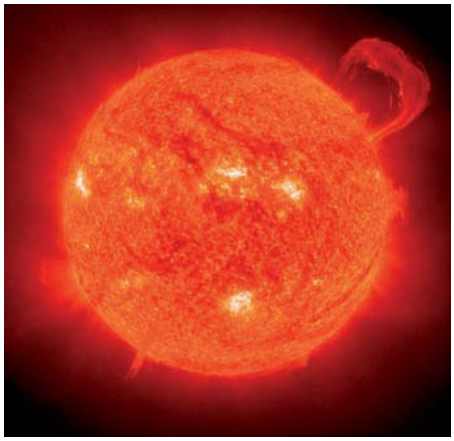


Figure 36.9 The Sun.

Class discussion

Since our oil and natural gas took millions of years to form and are being used up at such a rapid rate in our lifetime today, they are therefore finite energy resources. Which of coal, nuclear energy or renewable energy (e.g. solar or wind) would be the best energy solution for the Caribbean to be pursuing from now?'

In preparation for the discussion, you are encouraged to research some facts, such as the following:

- How much energy does the Caribbean need per year in order to function at a level in which its citizens enjoy a decent standard of living?
- What are some possible energy sources within the Caribbean?
- Will the sources be adequate to meet annual territorial or Caribbean-wide needs?
- How long is the energy source expected to last, given the current energy needs per year?
- How available is the energy likely to be on a continuous day-and-night and day-by-day basis?
- What are some environmental and public health implications?
- What are some disposal problems?

Chapter summary

- Energy may be obtained from the nucleus by fission or by fusion.
- In fission, a heavy nucleus, bombarded with a slow-moving neutron, breaks up into two fragments and extra neutrons are released.
- In fusion, light nuclei, travelling at high speeds, collide and fuse to form heavier nuclei.
- In both fission and fusion, some mass, Δm, disappears, having been converted into energy, $E = \Delta mc^2$, where c is the speed of light in vacuum.
- In a nuclear reactor, high-speed neutrons released during fission of nuclei initiate further fission reactions, when slowed down by a moderator.
- A self-sustaining chain reaction results when a certain minimum mass (called the critical mass) of fissionable material is present in the reactor core.
- Control rods of boron and cadmium, which absorb neutrons, help to keep the chain reaction from getting out of control.
- Overheating, with the possibility of explosions or meltdown, and high levels of radiation are two major dangers associated with nuclear power plants. In addition, radioactive waste produced by these plants and radioactive contamination present serious disposal problems.
- Nuclear energy obtained by fission could meet the world's energy demands for several centuries to come.
- Nuclear fusion, if it could be controlled, holds great promise as a major environmentally friendly, almost limitless, energy source for the world.

Answers to ITQs

ITQ1 Crude oil (40%)

ITQ2 Energy used per year = $5000\,\text{J} \times 3600 \times 24 \times 365$

No. of kg of fuel needed = $\frac{5000\,\text{J} \times 3600 \times 24 \times 365}{20 \times 10^6}$

$= 7884\,\text{kg}$

$\approx 8000\,\text{kg}$ (8 tonnes)

ITQ3 Steam must enter a turbine at high pressure and leave the turbine at low pressure in order for the turbine to spin efficiently. Since pressure is proportional to absolute temperature, the steam must be cooled as much as possible as it leaves the turbine. Water from a nearby lake, river or the sea is pumped in pipes and provides the required cooling.

ITQ4 Since nuclei are positively charged, it is difficult to get them to come together and fuse – the electrostatic force of repulsion becomes too large and they get close to each other. Temperatures of the order of a million kelvins are required to give the nuclei the high speeds needed to overcome the electrostatic repulsion. Designing containers for substances at such temperatures becomes problematic. These are only a few of the difficulties in obtaining controlled fusion power.

Examination-style questions

1 How much energy is produced when mater of mass of 1 g is entirely converted into energy?

A $3 \times 10^5\,\text{J}$ C $9 \times 10^{13}\,\text{J}$

B $3 \times 10^8\,\text{J}$ D $9 \times 10^{16}\,\text{J}$

2 Which would be most suitable to use as a coolant in a gas-cooled nuclear reactor?

A carbon dioxide C oxygen

B hydrogen D sulphur dioxide

3 Which of the following makes use of a fission reaction?

A nuclear power stations

B diesel power stations

C cobalt-60 in cancer treatment

D production of solar electric energy

4 How many neutrons are present in a $^{235}_{92}\text{U}$ nucleus?

A 92 C 235

B 143 D 327

5 The following equation illustrates the splitting of a uranium-235 nucleus in a nuclear reactor core:

$$^{235}_{92}\text{U} + {}^{1}_{0}\text{n} \rightarrow {}^{144}_{56}\text{Ba} + {}^{90}_{36}\text{Y} + 2\,{}^{1}_{0}\text{n} + \text{energy}$$

(i) The equation is not balanced. Balance the equation.

(ii) What is the name given to the splitting of the uranium-235 nucleus?

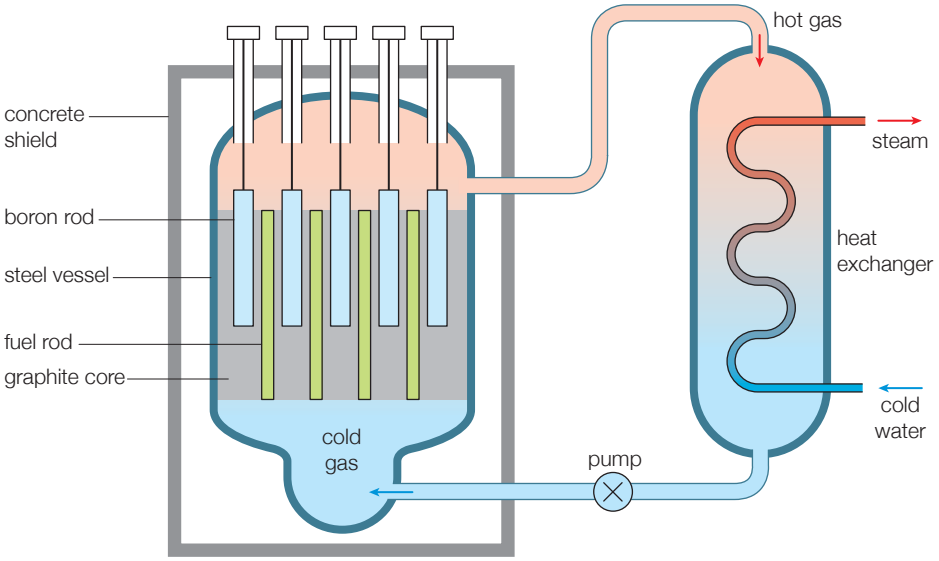
(iii) Name the element denoted by the symbol Y (not its correct chemical symbol!) (Use the Periodic Table, figure 33.3, if necessary.)

(iv) What feature of the equation:

(a) shows why a sustained chain reaction is possible?

(b) accounts for energy released during the reaction?

6 A simplified drawing of a nuclear reactor is shown in the diagram.



(i) State the function of the following:

(a) the boron rods;

(b) the graphite core;

(c) the fuel rods;

(d) the cold gas;

(e) the heat exchanger;

(f) the steam.

(ii) Describe *one* advantage and *one* disadvantage of using nuclear fission to produce electricity.

(iii) Describe a possible sequence of outcomes that could result if the pump shown in the diagram were to fail.

7 Give *one* argument in favour and *one* against for *each* of the following methods of disposal of radioactive waste stored in concrete containers:

(i) dumping in the deep sea;

(ii) burying deep underground;

(iii) sending into space.

8 The table shows the masses of a proton, a neutron and a helium nucleus. Calculate the mass 'lost' when 2 protons and 2 neutrons come together to form a helium nucleus.

	Masses of particles
proton	1.6726×10^{-27} kg
neutron	1.6749×10^{-27} kg
helium nucleus (2 protons + 2 neutrons)	6.6463×10^{-27} kg

Section F:

School-Based Assessment

37 School-Based Assessment

Practical work in physics

The present CSEC Physics syllabus (2013) makes clear that assessed practical work – the School-Based Assessment (SBA) – is an integral part of a student's studies. This aspect of the course gives the chance to personalise the curriculum to meet students' particular needs and assess the development of their skills.

Specified topics

In Physics, assessment in at least 16 exercises (spread across a minimum of 7 specified topics) is needed to satisfy the CXC requirements. The specified topics are:

1 Pendulum
2 Momentum/Conservation of Energy
3 Specific Heat Capacity/Specific Latent Heat Capacity
4 Refraction
5 Series and Parallel Circuits
6 *I–V* Relationships
7 Radioactivity Decay (Simulation)

This chapter includes outlines of 13 activities (in addition to those mentioned in the syllabus itself), which are suitable, after proper development, for use in SBA practical investigations. There are both qualitative and quantitative investigations. The chapter contains at least one practical exercise associated with each of the specified content areas, as well as other topics encountered in this course.

Sufficient detail is given to make possible the practical conduct of each experiment, and each one can be developed to illustrate material in the text. Each gives students the opportunity to develop their experimental and reasoning skills and also their ability to present results in the clear, appropriate way detailed in the syllabus.

Assessment of skills

If you are doing an experiment in class as part of your week's work, your teacher may have done some or all of the planning for you, collected all the materials that you need, and given you instructions how to do the work, or even a written worksheet to follow. But you still have to show that you can follow the instructions, do the experiment and present your results well. The skills which will be tested are:

Experimental skills (X/S)

- Manipulation and measurement (M/M)
- Observation, recording and reporting (O/R/R)
- Planning and designing (P/D)

Use of knowledge (U/K)

- Analysis and interpretation (A/I)

Some background to practical work: what was Newton's contribution to scientific methodology?

Not much was known about what is now known as the 'scientific method' before Newton used it in his experiments to examine colour in the late 17th century. Until this time, if a scientific idea was to be tested or a phenomenon was to be investigated, experiments were conducted, but not in a structured manner. Factors that might have had an effect on the matter being investigated were not examined one at a time while the others were held constant. In other words, not much was known about 'controlling variables' in experimental work.

Newton was keen to find out why images obtained in telescopes (which had then just been discovered) were often coloured, and so he set himself the task of studying the composition of white light. He also investigated the formation of the light spectrum and whether the colours of the spectrum could be put together again to produce white light. In all his experiments, Newton was careful to investigate the effect of only one factor at a time.

A three-step approach in preparing for your SBA

When planning and presenting your project, your SBA has three parts:

1 Planning and designing the experiment.

2 Doing the laboratory work.

3 Presenting a lab report.

1 Planning and designing the experiment

In any experiment, you are trying to find an answer: it might be a relationship or a value. You will need to devise and follow a logical series of steps to find out that information. Therefore, whatever your hypothesis, you must have a plan.

In some cases, the proposed activities already in this section contain an outline plan. However, you would still need to design your experiment bearing in mind the equipment and facilities available in your lab. In other activities, you are given a problem to solve, and here you would have to plan and design your investigation from the beginning. In this latter case your work would most likely be assessed under planning and designing [P/D].

Part of your planning is specifying, in detail, how to carry out your experiment. You need to plan the experiment in such specific detail that someone else reading your design would be able to do exactly the same as you did and get the same results.

Think about:

- What apparatus will you need? E.g. 'a 250 cm^3 beaker' (and not just 'a beaker').
- What chemicals will you need? E.g. '3 or 4 potassium manganate (VII) crystals about 2 mm long'.

- What will you do? E.g. 'stir the mixture gently with a sturdy plastic drinking straw just before taking each temperature' (not just 'stir the mixture and take the temperature').
- What could go wrong?
- What are some possible hazards? What safety precautions should be taken?

You will need to put your plan in writing. It is always a good thing to have your teacher, as well as classmates, check your plan before attempting to carry out your investigation.

Often, no matter how carefully you have planned an experiment, it doesn't go as you thought it would. However, it has still told you something – an experiment never 'fails'. If it didn't produce the effect you needed, find out why. If the experiment 'worked' but didn't give the result you expected then you've found out something new.

2 Doing your laboratory work

Here you carry out your plan. Always consult with your teacher if you are not sure exactly how to use the equipment, and how to use it safely. CSEC expects you to have practised using equipment before being formally assessed in their use.

3 Presenting the lab report

Lab reports should be written up using the following format.

Writing in the correct tense
Remember that your research is already finished. Use the past tense when talking about the experiment: 'The objective of the experiment was ...' or 'The mixture was added to the beaker'.
Your report, the theory you are testing and your equipment still exist; therefore, these get the present tense: 'The purpose of this report is ...' .

Date and Title

The title should be brief. It should describe the main point of the experiment or investigation.

Aim

Keeping it simple and achievable, describe the purpose of the experiment.

Discussion

Background information to relate the experiment to something you have learnt or seen, or to a problem you have faced.

Hypothesis

Should be a clear statement proposing an answer to a question that you want to find an answer to. For example, 'Vitamin C in orange juice oxidises over time when exposed to air.'

(From the data you obtain from your experiment you should be able to say if the hypothesis has been either supported or not in your conclusion.)

Procedure

Use the past tense (see above). In a bulleted list or in separate paragraphs, state, in order, what you did. Include clear, quantifiable detail (e.g. quantities stated and apparatus specified). Your teacher will suggest that you use either of these two forms of words:

'I washed a 250 cm^3 beaker ...' or 'A 250 cm^3 beaker was washed ...'

If you have written a good account then someone else, having read it, should be able to repeat the experiment exactly as you did it without any other help.

Diagram

Draw the apparatus as neatly as you can.

Results

- *Table – title stated, neatly drawn with accurate data (times, volumes, masses, colours ...) and proper units for quantities.*
- *Graph (if necessary) – title stated, axes labelled with proper units, points accurately plotted. Use a line or a bar graph as required. Remember that you should choose the correct type of graph for the data you are presenting.*
- *Include any calculations you used.*

Explain the results in detail, using values found in your results

Answer the questions given in paragraph style.

Limitations

Explain any condition or factor that is out of your control and which affects the results obtained.

Conclusion

One short paragraph to summarise results. Make it related to the aim. Review your data and state your opinions and arguments of what the results show, for example: 'As graph 3 shows, there is a marked difference between group A and group B, which allows the conclusion that .....' .

State if your hypothesis has been supported or not. If the data you have obtained is not sufficient to support or reject the hypothesis, state why and say what further work could be done that would allow you to draw a stronger conclusion.

If you are undertaking a complete project (which will be the case in the second year when you carry out your investigation) then more will be expected of you, and you can see from page 55 of the syllabus how marks for the project will be awarded. Your report will need to be more comprehensive than for a class experiment. It will be assessed for planning and designing and for analysis and interpretation. Planning and designing has twice the marks of analysis and interpretation.

A sample investigative write-up for SBA

Name:

Grade/form:

Date:

Resistance of a wire

title

Part A – The proposal

Observation

You are given a roll of insulated resistance wire. Unfortunately the label with the specifications is missing. You require several resistances of 20 Ω each for an experiment and need to know what length of the wire would give this resistance. Since your supply of resistance wire is limited, you are only allowed to cut off about 30 cm for your investigation.

background information gives context to the experiment

Design an experiment so that you can know exactly what length you will need for a 20 Ω resistance.

your task

Hypothesis

50 cm of the resistance wire will have a resistance of 20 Ω.

Aim

To determine the length of a given insulated resistance wire needed to construct a 20 Ω resistance, using Ohm's law and a graphical method.

the aim describes the purpose of the experiment in no more than two or three sentences

Apparatus

this is another word for 'equipment'

Resistance wire; wire cutter and stripper; ammeter 0–1 A (× 0.02 A); voltmeter 0–5 V (× 0.02 V); 0–6 V power pack; variable resistance (or rheostat) 0–10 Ω; switch or key; connecting wire; alligator clips (to attach the test wire to the circuit); about 10 cm masking tape; wooden strip, e.g. wooden 30 cm ruler; metre rule.

Circuit diagram

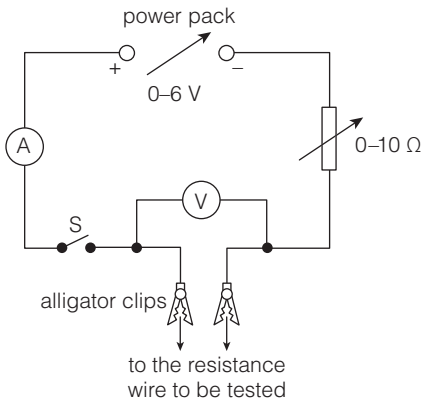


Figure 37.1 Circuit diagram for the determination of the resistance of a length of wire.

Variables

Independent – current in resistance wire (I)

Dependent – voltage across resistance wire (V)

Controlled/constant – temperature of resistance wire

Method

1 Wind the wire around the wooden strip and use the masking tape to keep it in place on the strip.
2 Attach the alligator clips to the ends of the wire and immerse the whole wire in a beaker of melting ice.
3 Refer to the circuit diagram for instructions on how to complete the setup.
4 Set the rheostat at its highest value of resistance with the switch off and with the power pack set at the lowest setting, 1 volt.
5 Slowly reduce the rheostat resistance, taking due care that the current and voltage readings keep within the ranges of the ammeter and the voltmeter.
6 Record the maximum current value in the table and the corresponding value of the voltage across the wire shown by the voltmeter.
7 Begin with the lowest rheostat resistance (highest current and highest voltage for the test resistor) and reduce the current flowing in the circuit by means of the rheostat.
8 Record (in a table) the values of the current at the pre-set voltages. Do this for each current at a pre-set voltage until the lowest current is reached.
9 Reverse the procedure and record values of current and voltage with the current increasing once again.
10 Plot the corresponding values of current and voltage on a graph using suitable scales.
11 Draw a line of best fit.

Expected results

The graph of voltage against current is expected to be a straight line, passing through the origin, since, according to Ohm's law, voltage is directly proportional to current in a wire (providing that the temperature remains constant). The slope of the graph would be the resistance of the 30 cm wire. The resistance per unit length of the wire can be calculated knowing the resistance and length – hence the length needed for a 20 Ω resistance can be determined.

What are the key points to note in the model proposal?

- The aim and the theoretical principle to be used (Ohm's Law) of the experiment were clearly stated.
- The future tense was used.
- The 'voice' used was the passive voice. The subject of sentences was never 'I'. Remember that you never mention yourself in writing reports.
- Care was taken to mention the steps in the order in which they were undertaken.
- A line diagram (and not a perspective one) of the arrangement of the apparatus used was supplied and the accepted symbols were used to 'label' the components so they could be recognised.
- The diagram was given a general caption.
- An attempt was made to be precise in order to keep the account as brief as possible, but, at the same time, complete.
- No values of readings taken of the variables were mentioned in the method. All readings would have been given in the table in the Implementation section.

Find out from your teacher whether you might be able to carry out this activity as a group. It is an excellent experiment to try for all-round experience in dealing with electrical circuits.

Part B – The implementation

Method

1 The test wire was first wound around the wooden strip and the tape wrapped over it to keep it in place.

2 The alligator clips were then attached to the bared ends of the wire and the entire length of the wire immersed in a beaker of crushed melting ice.

3 The rest of the circuit was set up as shown in the circuit diagram with the rheostat set at its highest value of resistance, the switch off and with the power pack set at the lowest setting, 1 volt.

4 A test run of the experiment was first conducted to ensure that the value of the applied voltage of the power pack was suitable for the values of the components in the circuit,

5 The test wire was connected in series with the rest of the circuit, the switch closed and the readings of the ammeter and the voltmeter recorded on the table shown below. These readings of current and voltage were the lowest values that would be used.

6 The rheostat resistance was then slowly reduced and, as the value of the current in the circuit rose, this rise was monitored closely until the ammeter reading reached full-scale. This maximum current value was recorded on the table, as was the corresponding value of the voltage across the wire shown by the voltmeter. These readings of current and voltage were the largest values that would be used in the experiment. The current was then switched off.

7 The experiment proper was now started. The range of voltage to be used was divided up into five convenient and roughly equal intervals and the voltages defining these intervals recorded on the table.

8 Axes were also set up on graph paper (since the range of values of V, the voltage, and I, the current, were known) such that the graph would occupy more than one half of the graph sheet. The axes started from 0, 0.

9 Starting with the lowest rheostat resistance (highest current and highest voltage for the test resistor) the current flowing in the circuit was reduced

by means of the rheostat, and the values of the current at the pre-set voltages were read and recorded on the table at these voltages.

10 This was done for each current at a pre-set voltage until the lowest current was reached. Then the procedure was reversed and values of current and voltage were recorded with the current increasing once again.

11 Each time there were corresponding values of current and voltage the corresponding point was plotted.

12 The line of best fit was then drawn among the points on the graph paper so that, judging by eye, the total positive error balanced out the total negative error.

Results

Table 37.1 Results for current and voltage.

	Current in wire, *I*/A			Voltage across wire, *V*/V		
	1st	2nd	mean	1st	2nd	mean
1	0.00	0.00	0.00	0.00	0.00	0.00
2	0.20	0.20	0.20	0.70	0.70	0.70
3	0.40	0.40	0.40	1.30	1.20	1.25
4	0.60	0.60	0.60	1.80	1.90	1.85
5	0.80	0.80	0.80	2.40	2.40	2.40
6	1.00	1.00	1.00	3.00	3.10	3.05

Graph

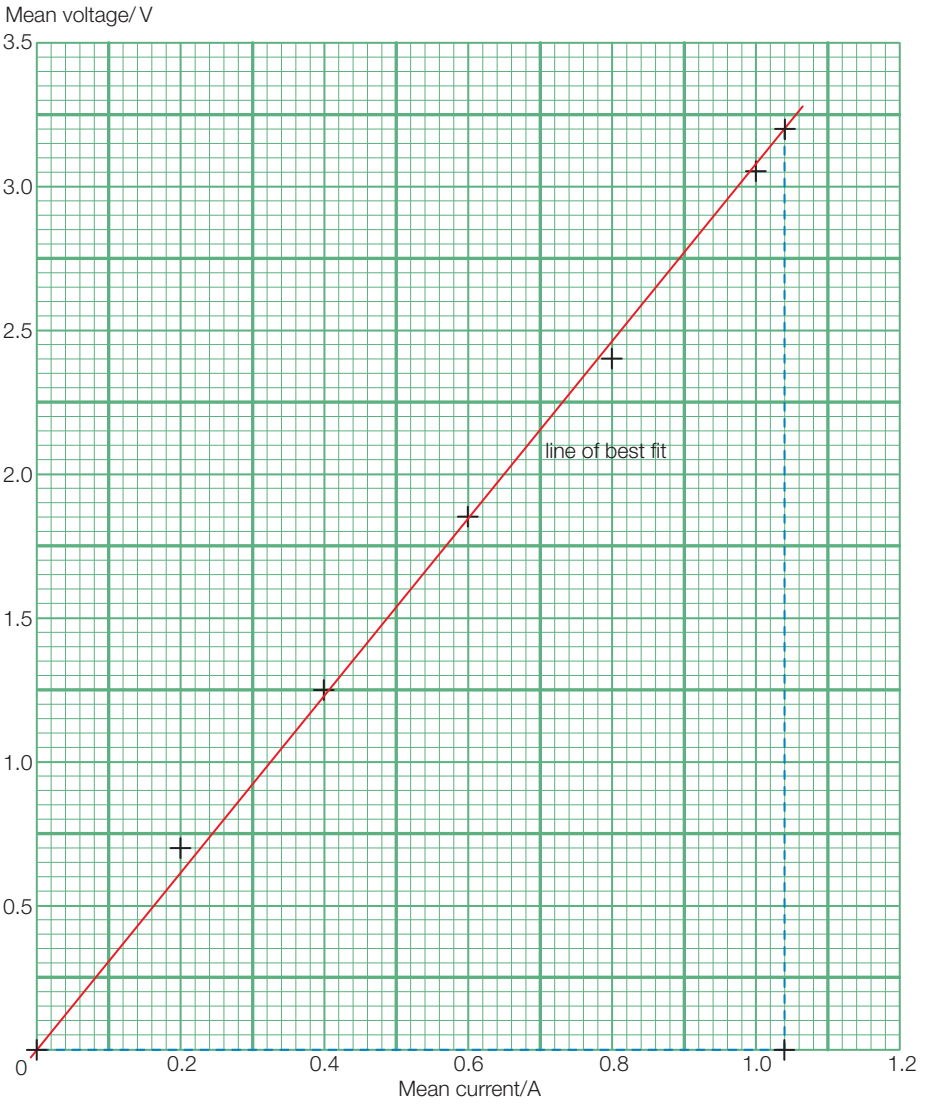


Figure 37.2 A graph of the results from table 37.1.

Calculations

(a) Since the graph passes through the origin (0, 0), then its equation must be

V, the voltage across the wire

= (a constant, K) × the current, I, in the wire

or $V = k \times I$, where k = the slope of the graph.

The relation between the variables is, therefore that the voltage varies directly as the current

or $V \propto I$

Since

$V = k \times I$

Then the ratio $\frac{V}{I} = k$

But by Ohm's Law $V = R \times I$ (where R is the resistance of the wire)

and so $R = \frac{V}{I} = k$(the slope of the graph)

The slope of the graph, $k = \frac{(12.80\,\text{cm} \times 0.25\,\text{V cm}^{-1})}{(10.40\,\text{cm} \times 0.10\,\text{A cm}^{-1})}$

$= \frac{(3.20\,\text{V})}{(1.04\,\text{A})}$

$= 3.076\ldots\ \text{V/A}$

$= 3.1\,\text{ohms, since } 1\,\text{V}/1\,\text{A} = 1\,\text{ohm}$

(b) Since 30 cm of wire has a resistance of 3.1 Ω, then length needed for 20 Ω

$= \frac{30\,\text{cm} \times 20\,\text{ohms}}{3.076\,\text{ohms}}$

$= 195\,\text{cm}$

Note that the slope of the graph is calculated with the units of the quantities giving the slope.

Important points to note

1 Since the graph shows no intercept, then the relationship is a direct variation.

2 The graph occupies a large fraction of the graph sheet and this allows a large enough scale to make plotting errors small.

3 The corners of the triangle used for taking the slope are so chosen that their co-ordinates are easy to read off,

4 Both the numerator and the denominator used for obtaining the slope are expressed with their units.

5 We have used two (2) significant figures for the resistance since the smallest values of the voltage and the current shown in the table each have two significant figures, as we should 'follow suit' and do the same, though some might argue that since both of the quantities being divided to give the resistance are expressed to three figures we should do the same and use three figures.

Discussion

Some of the readings were not steady as these might have been due to voltage fluctuations of the power source or varying contact resistance. There might have also been some heating of the wire due to current in it, causing the resistance of the wire to change. Further, the resistance of the voltmeter in particular could have contributed to errors in voltage readings. Both the ammeter and the voltmeter also have inherent manufacturer's percentage accuracy ratings. Added to these, there were errors that could not be avoided, such as those due to the precision of the instruments. The value obtained for R must therefore be interpreted, taking all these factors into consideration.

Limitations

If a dry cell is used to conduct this experiment, there would most likely be 'voltage drop' and current drift. If this occurs, use a lead battery.

Reflections

Suppose the graph of V against I was a 'good' straight line, in that most of the pints were on the line, but it showed an intercept. What conclusions would you draw?

Conclusion

The value of the resistance of 1 metre of the wire under test = …, and so the length of wire needed to make a resistor of resistance of 20 ohms = … m.

Safety first

There are several sources of danger that you need to address as you develop your SBA activities. There are dangers to yourself, your classmates, the equipment and even the school building itself. Here are a few safety symbols concerning situations you should bear in mind:

					
Electrical hazard	Hot surface	Laser light	No food and drink allowed	Corrosive substance	Flammable materials
					
Biohazard	Wear safety goggles	Radioactive	No pointed objects	Toxic materials	

School-Based Assessment contents

9.1 To find the efficiency of a block-and-tackle pulley system 564

13.1 Comparing the specific heat capacity of salt water by the method of mixtures 566

13.2 Determining the specific heat capacity of milk (electrical method) 567

13.3 Investigating the identity of a metal solid through its specific heat capacity (method of mixtures) 568

22.1 To find the refractive index of Perspex in air 569

23.1 To find the focal length of a converging lens 571

23.2 To locate the image formed by a glass prism 573

26.1 Investigating the current–voltage relationship in an ionic solution 576

26.2 How much current could an LED safely handle? 577

26.3 Series and parallel circuits 578

28.1 The efficiency of energy conversion in a d.c. motor 579

35.1 Simulating radioactive decay 580

35.2 Investigating radioactivity in food 581

Note

- All the experiments that follow might not have been presented using the same format. However, in writing up your lab reports, you are urged to use the CSEC® reporting format to encourage good habits in your coursework.
- The skills listed at the beginning of each activity are meant only to indicate possible areas of SBA. It is not likely that you will be assessed in more than two skills in any one experiment.

© Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014 **Photocopiable**

9.1 To find the efficiency of a block-and-tackle pulley system [M/M]; [O/R/R]; [A/I]

Syllabus reference A5

Introduction

The efficiency of a machine is given by

$$\text{efficiency} = \frac{\text{energy usefully converted by the machine}}{\text{total energy supplied to the machine}} \times 100\%$$

$$= \frac{\text{work done to raise the load}}{\text{work done by the effort in doing so}} \times 100\%$$

and we also have

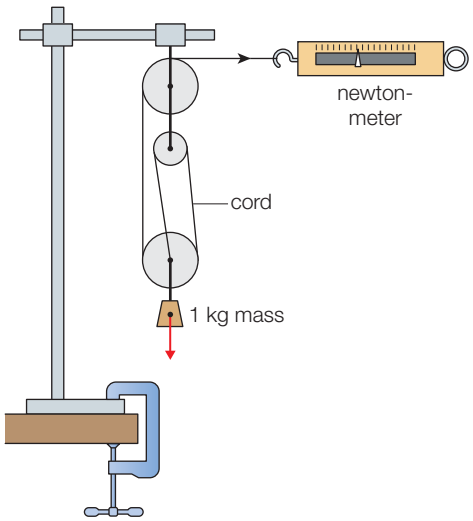
work done = force × distance moved by the force in the direction of the force

Apparatus

Block-and-tackle pulley system (preferably three pulleys); retort stand and clamps; G-clamp; length of light cord; metre rule; 1.0 kg mass (producing a load *W* of 10 N).

Procedure

1. Set up the block-and-tackle system on a retort stand as shown below. Make sure that the base of the stand is kept *firmly clamped* to the laboratory bench with a G-clamp.
2. Pass the cord over and under the pulleys. Attached the end that was passed round the pulleys to the lower block. Tie the 1.0 kg mass to the system as shown in the diagram. Put something soft on the floor under the mass – just in case.
3. Attach a newton-meter to the free end of the cord.
4. Pull the newton-meter, with the free end of the string horizontal, through a measured distance (*x*) of about 1 metre. While you do this, make sure that the reading of the newton-meter (*P*) remains constant. Measure the distance (*y*) through which the load *W* rises. Record the readings *P* (in newtons) and *x* and *y* (in metres) in a table like the one shown.



Photocopiable © Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014

5 Repeat the measurements of P, x and y for the same value of x (1 m) twice more. Record these readings.

6 Then repeat the experiment using values of x of 90 cm, 80 m, 70 cm, 60 cm and 50 cm. For each of these measurements, measure the distance through which the load W rises.

load (constant) $W = 10$ N							
effort distance		effort		load distance		Px/J	Wy/J
x/cm	mean	P/N	mean	y/cm	mean		
100, 99.0, 100	99.7			33.0, 37.0, 33.0	33.0		
91.0, 90.0, 92.0	91.0			30.6, 30.0, 31.0	30.5		

Handling your results

1 Draw a diagram of the apparatus.

2 You know, for each experiment, the load, the effort used, the distance moved by the load, and the distance moved by the effort. Record your results in a table like that shown (which is not complete).

3 Calculate, for each distance through which the effort moved, the work done by the machine and the work input by the effort.

4 Draw a graph of 'work done to raise the load' against 'work done by the effort'.

5 From the graph, calculate the ratio $\frac{\text{work done to raise the load}}{\text{work done by the effort in doing so}}$.

6 Calculate the efficiency of the machine.

Questions

1 Why did repeating each experiment increase the accuracy of your results?

2 What shape of graph do you expect? Why?

3 How will you obtain the ratio $\frac{\text{work output}}{\text{work input}}$ from your graph?

4 Why is the efficiency of the system *not* 100%?

5 How could you make the experiment more accurate?

6 Do you think the efficiency would be the same for various loads? Explain your answer.

© Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014 **Photocopiable**

13.1 Comparing the specific heat capacities of salt water and fresh water using the method of mixtures [P/D]; [O/R/R]; [A/I]

Chapter 13 Measuring heat energy

Syllabus reference B3

Introduction



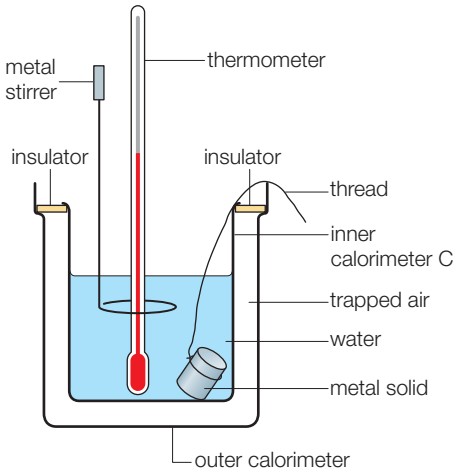
The natural waters found in and around Guyana are basically 'brackish' (i.e. have a low salt content). This is due to the effects of the great Amazon River, as well as the three large rivers in Guyana, pouring out their huge amounts of 'fresh' waters into the Atlantic Ocean on to the north coast of South America. However, most of the Caribbean territories are surrounded by the very salty water of the Atlantic Ocean.

Is there a difference between the specific heat capacities of water that is fresh and water that is salty? You are to design and carry out an experiment to determine if there is much difference between the specific heat capacities of fresh water (e.g. rain water) and salt water.

(Hint: An apparatus arrangement similar to the figure below can be considered.)

Proposal

Discuss your investigation with colleagues and write up a proposal (see page 556). Discuss your written proposal with your teacher.



Investigating and reporting

After your proposal has been approved by your teacher, carry out the investigation and write up a Lab report (page 554). In the Reflections section, you may wish to comment on:

(a) the size of the difference (if any) between the specific heat capacities;

(b) the implications of (a) as regards temperature variations in the two types of territories;

(c) whether one pair of values of specific heat capacities is enough in order for a reasonable conclusion to be made.

Photocopiable © Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014

13.2 Determining the specific heat capacity of milk (electrical method) [P/D]; [O/R/R]; [A/I]

Chapter 13 Measuring heat energy

Syllabus reference B3

Introduction



The energy required for refrigeration of milk depends partly on its specific heat capacity. You are to design an experiment to compare the specific heat capacities of two types of milk (e.g. whole milk and 2% milk) using an electrical method.

Proposal



Practical activity 13.1 (page 201) outlines a possible way of using an electrical method in determining the specific heat capacity of a liquid. Worked example 13.2 (page 202) shows how a calculation can be made. In your design, you are to indicate what you plan to do concerning safety considerations and in order to maximise the accuracy of your experiment.

Discuss your written plan and design with your classmates and with your teacher. Once your teacher has approved of your plan, you are to go ahead with the practical activity, bearing in mind safety precautions.

Investigating and reporting

After your proposal has been approved by your teacher, carry out the investigation and write up a Lab report (page 554). In the Reflections section, you may wish to comment on:

(a) the size of the difference (if any) between the specific heat capacities;

(b) the implications of (a) as regards energy required for refrigeration of the two types of milk, starting from room temperature;

(c) whether one pair of values of specific heat capacities is enough in order for a reasonable conclusion to be made.

© Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014 **Photocopiable**

13.3 Investigating the identity of a metal solid through its specific heat capacity (method of mixtures) [P/D]; [O/R/R]; [A/I]

Chapter 13 Measuring heat energy

Syllabus reference B3

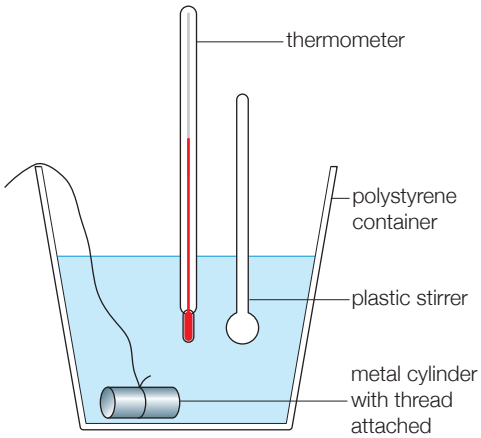
Introduction



There are various ways of checking what metals things are made of experimentally coins, for instance, are often made of nickel. Nails can be made of iron. In this exercise, you are to design and carry out an investigation using a method of mixtures experiment in an attempt to ascertain what your coin or nail is actually made from.

Proposal

With a little bit of thought, the arrangement in figure 13.11 (reproduced below) and the procedure associated with it (page 203), can be adapted in your design. (In particular, one design problem that you will face is that a coin, or a nail, for that matter, has such a small mass.)



Worked example 13.3 (page 203) shows how a calculation can be made. In your design, you are to indicate what you plan to do concerning safety considerations and in order to maximise the accuracy of your experiment. Discuss your written proposal with your classmates and with your teacher.

Investigating and reporting

Once your teacher has approved of your plan, you are to go ahead with the practical activity, bearing in mind safety precautions. In the Reflections section, you may wish to comment on:

(a) the identity of the metal, in relation to the value you obtained for its specific heat capacity;

(b) one other approach, not involving heat, that can be used to assist in validating either of the statements in the introduction.

Photocopiable © Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014

22.1 To find the refractive index of Perspex in air [M/M]; [O/R/R]; [A/I]

Chapter 22 Refraction of light

Syllabus reference C4

Introduction

When a ray of light passes from one medium to another, the refractive index of the second medium with respect to the first is given by the ratio $\frac{\sin i}{\sin R}$, where i and R are the angles of incidence and refraction respectively. The relative refractive index is a constant for a particular pair of media.

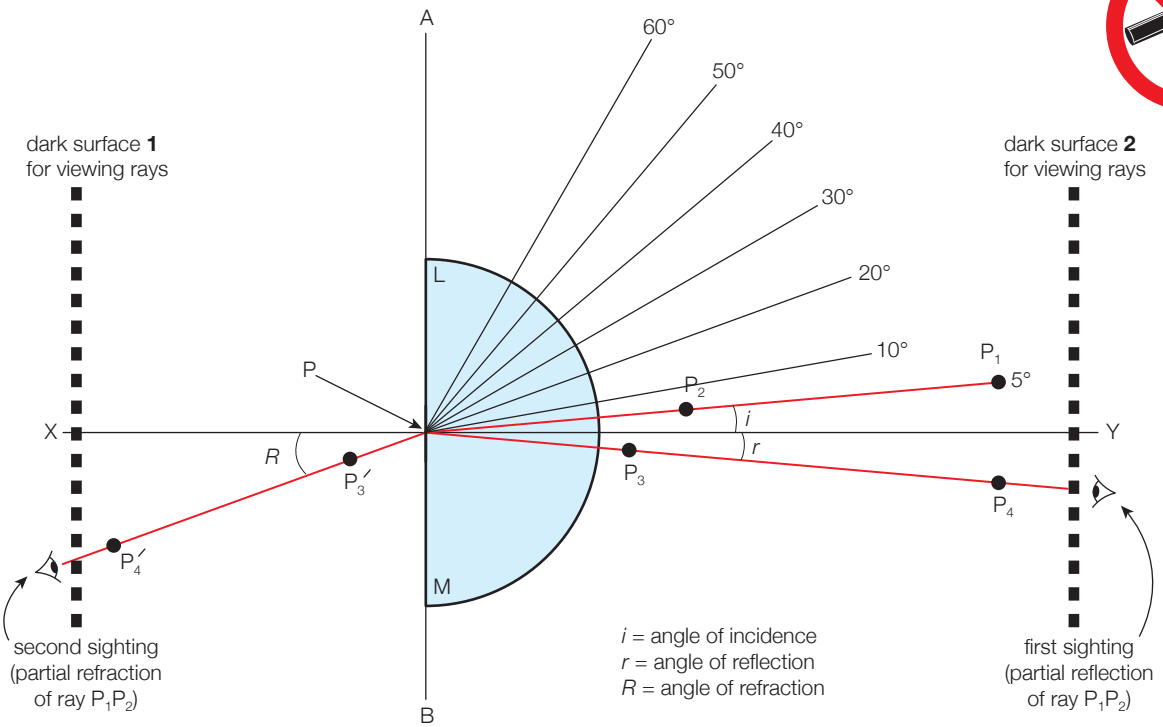
Apparatus

Semicircular Perspex block; protractor; sheet of A4 drawing paper; drawing board; four optical pins.

Procedure

1. In the middle of the A4 sheet of paper, construct two lines XY and AB so that they roughly bisect each other and are perpendicular. Let the lines intersect at P.
2. At P, construct angles of 5°, 10°, 20°, 30°, 40°, 50° and 60°, respectively, with XY. Mark the angles accordingly. These angles will be the angles of incidence i.
3. Mark the centre of the edge LM of the block. Place the semicircular block so that this edge is directly above AB, and with the marked centre of LM directly above the point P (as shown in the diagram).
4. Stick two pins, P_1 and P_2, on the 5° line about 10 cm apart.





5. With one eye closed, find the direction along which the images of P_1 and P_2, seen from the curved side of the block, are in line. Stick two more pins (P_3 and P_4 on the diagram) into the paper so as to be in line with these two images. (The use of the dark surface 1 will help to make the images of the pins easier to see.)

© Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014 **Photocopiable**

6. Remove the pins P_3 and P_4, mark their positions accordingly and join these positions with a straight line. This line represents the path of the incident ray P_1P_2 when reflected at the point P. This image will be formed by partial reflection.
7. Now turn the drawing board round (or move to the other side of it) and look from the side of the edge LM of the block to find the direction in which the pins P_1P_2 again seem to be in line from refraction at the point P. Stick the two pins (now P_3' and P_4' on the diagram) into the paper about 10 cm apart to mark this direction. Use the dark surface 2 for this sighting.
8. Remove these two pins and mark their positions by P_3' and P_4'. Join these two positions with another straight line.
9. Remove the block and measure the angle of reflection *r* and the angle of refraction *R*. Record your results in a table.
10. Repeat steps 3–9 for the other angles marked out.

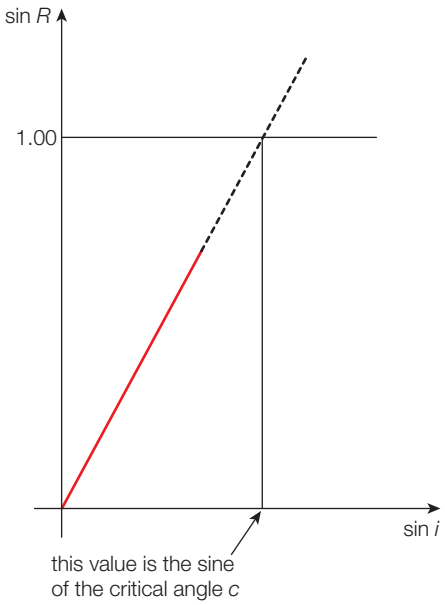
Handling data

1 Check to see how closely $i = r$, for reflection, for the values of *i* used.

2 Plot a graph of sin *R* against sin *i*, for refraction. Find the refractive index, sin *i*/sin *R*, from the graph.

3 Extrapolate the graph to find the value of sin *i* for which sin *R* = 1.00. Hence find the value of *i*. Denote this angle by *c*, since it is the critical angle for the Perspex–air interface.





Questions

1 Why is this general method of finding the critical angle better than placing the two pins by trial-and-error until you can no longer see their refracted image?

2 For which of the values of *i* used were you able to get partial reflection? Why did you not get partial reflection for the others?

3 Were the images obtained by partial reflection brighter or fainter than those obtained by total reflection? Give a reason for what you observed.

4 What do you think would happen to a ray if it were to travel to P from the curved side of the block along the line XY – would it suffer both reflection and refraction?

5 How would you use the apparatus to locate the position of the image of a pin placed on the line XY on the curved side of the block?

Photocopiable © Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014

23.1 To find the focal length of a converging lens [M/M]; [O/R/R]; [A/I]

Chapter 23 Lenses

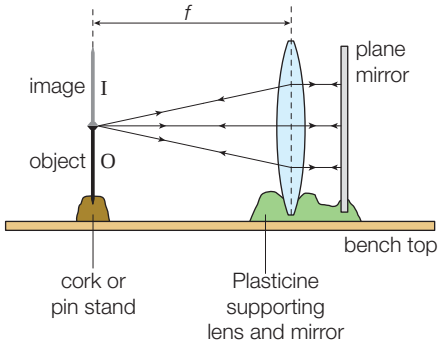
Syllabus reference C5

Introduction



A convex lens is placed in front of a mirror and a pin is placed on its principal axis. The position of the pin is changed until it is coincident with its own image. For this to happen, the rays leaving the pin must strike the lens and the mirror behind it and return along the same path, as shown. This in turn can only happen if the rays strike the plane mirror perpendicularly. Then, the rays must be parallel to the axis, since the plane mirror is perpendicular to the axis. Finally, rays can only travel parallel to the axis if they came from the principal focus of the lens.

Here we are using the reversibility principle, that is, if rays travelling parallel to the axis meet at the principal focus after passing through the lens, then the reverse will be true. If they reverse their path and travel parallel to one another on their way to the lens, they will all meet at the principal focus after refraction.

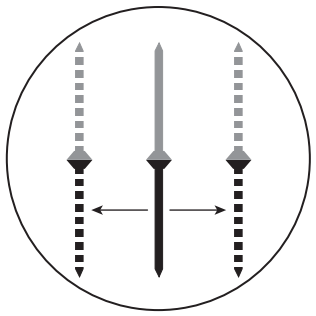


Apparatus

Converging lens; lump of Plasticine or a lens stand; plane mirror; one optical pin or a small nail; a cork (or pin stand) to support the pin.

Procedure

1. Support the lens in a vertical plane on the lump of Plasticine or in the lens stand.
2. Support the plane mirror behind the lens so that its reflecting surface is parallel to and facing the lens. The lens and mirror can either touch each other or be separated by a distance of up to about 2 cm.
3. Place the pin (or nail) in front of the lens so that it is roughly on the axis of the lens, as far as you can judge.
4. With one eye closed, adjust the position of the pin until the image of the pin and the pin itself are both in the middle of the mirror. This is the starting point of the experiment.
5. Check for no-parallax between the pin and its image as seen in the lens–mirror system. At the end-point, the two should move together for at least half of the width of the lens (as shown).
6. If no-parallax is not obtained, adjust the pin position for initial alignment as at step 4 and again check for no-parallax.
7. Repeat the check for alignment and the test for no-parallax until the end-result, no-parallax, is obtained, as shown. This is the end-point of the experiment.
8. Measure the distance between the pin and the centre of the lens. This distance is the focal length of the lens.
9. Repeat steps 3–8 at least three more times, then find the mean value of those obtained. This value is the focal length of the lens.



no parallax

Questions

1. The measurement made in step 8 of the procedure is not easy to do accurately. How can you overcome this problem?

© Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014 **Photocopiable**

2 How could you make an experiment to test the accuracy of the method of no-parallax? Does it detect whether the pin and image are 1 cm apart, 1 mm apart or less? If 'yes', explain how.

3 How does this affect the accuracy of your result for the focal length of the lens? Estimate the precision of your result.

 Photocopiable © Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014

23.2 To locate the image formed by a glass prism [M/M]; [O/R/R]; [A/I]

Chapter 23 Lenses

Syllabus reference C5

Introduction

The path of a ray of light from an object can be traced by placing two optical pins so that they seem, from a certain direction, to be in line with the object. Any two (or more) such paths, extended until they meet, define the position of the image of the object.

Apparatus

Right-angled isosceles glass prism; four optical pins; large sheet of paper (A4 paper should be large enough); drawing board; drawing pins or sticky tape.

Procedure

1 Use sticky tape (or the drawing pins) to fix the sheet of paper to the drawing board. Draw a long line, XY, about half of the way down the sheet. With your protractor, and about 5 cm from X, draw a line ST perpendicular to XY.

2 Place the prism on one of its triangular faces so that the hypotenuse face AB is directly over the line XY and the right angled edge is away from you and on the other side of XY, as shown in the diagram.

3 Place a pin, P_1, on ST about 12 cm from the prism and XY. Place another pin, P_2, on the line ST about 2 cm from the prism and XY.

4 Close one eye and take the other eye down to the level of the pins.

5 Place your head in that position where the images of P_1 and P_2 are one behind the other, as seen in the hypotenuse face AB. When this is achieved, stick the other two pins in to the paper (P_3 and P_4 on the diagram) so as to be in line with the first two as seen in the hypotenuse face of the prism.

6 Draw circles round the points of pins P_1 and P_2 and round the points of pins P_3 and P_4. Mark the points P_1 to P_4 accordingly. Remove the pins P_2, P_3 and P_4. Do not remove pin P_1.

7 Now stick the pin P_2 about 2 mm to the right of its original position, at P_2'.

8 Look for the images of pins P_1 and P_2' in line as before. Stick the other two pins in such positions as to be in line with the images of P_1 and P_2' as seen in the prism (now P_3' and P_4' on the diagram). When this is achieved, draw circles round the three pin-points and mark them P_2', P_3' and P_4', respectively.

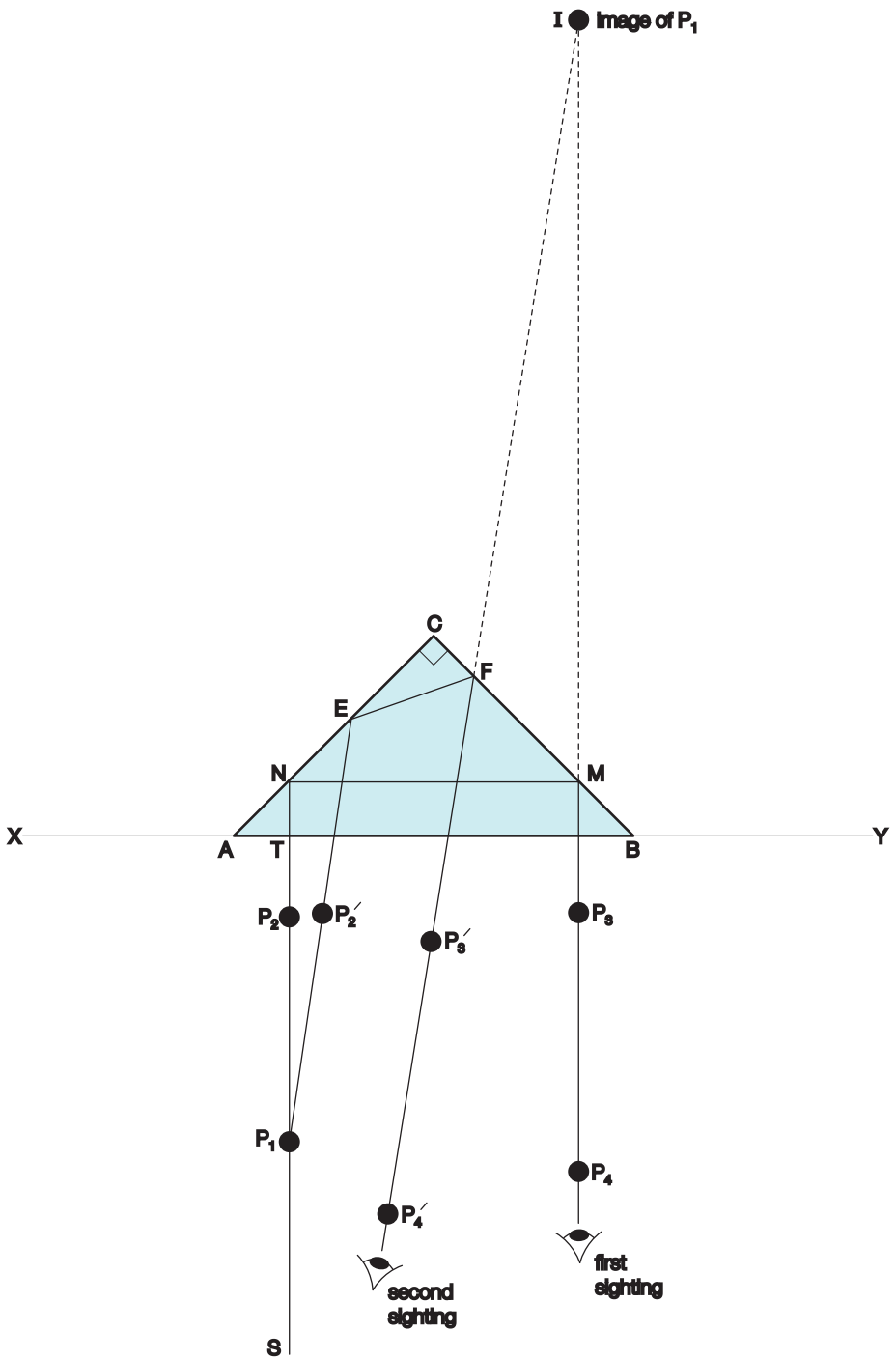
9 Draw in the outline of the prism and then remove the prism from the paper.

10 Produce P_1P_2 to meet the face AC of the prism at N and produce P_4P_3 to meet the face CB at M. Join NM. This line is the path of the ray P_1P_2 across the prism.

11 Join P_1P_2' and $P_4'P_3'$ and extend them to meet the faces AC and CB at E and F.

12 Produce P_4P_3 and $P_4'P_3'$ with broken lines to meet behind the prism. Label the point where they meet at I. The fact that the lines converge to a point shows that an image exists at that point.

© Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014 **Photocopiable**



 Photocopiable © Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014

Information

Using the usual notation:

$$\frac{\sin(\text{angle in air})}{\sin(\text{angle in medium})} = n_{\text{medium}}$$

these being the angles made by the ray with the normal.

Theory

If the refractive index of the glass of the prism is 1.55, then the critical angle c for the glass–air interface is given by

$$\sin c = \frac{1.00}{1.55} = 0.65 \quad \text{or} \quad c = 40.5°$$

Now the ray P_1P_2 will be totally internally reflected at the face AC, since the angle of incidence at this face is 45° ($>c$). The same will happen to the ray when it strikes the face BC, for the same reason. Hence the behaviour of the ray P_3P_4.

Much the same will happen to the ray P_1P_2', since it will cross the face AB of the prism only slightly deviated (towards the normal) from its original path. It will eventually emerge from the face AB as shown (along $P_3'P_4'$), but diverging from P_3P_4.

Interpretation of the results

Since the rays P_3P_4 and $P_3'P_4'$ are diverging when they leave the prism, this indicates the presence of a virtual image. When these rays are produced backwards, the point where they intersect is the position of this virtual image.

Questions

1 How would you try to verify that the image of the pin P_1 was indeed at the point, I, which you found in your experiment?

2 The diagram assumes that the ray P_1P_2 passes through the face AB without refraction. Why is this a correct assumption? Use the refraction law to show this.

3 The diagram assumes that the ray P_1P_2' passes through the face AB without refraction, but this is not true. Why is it untrue? Does the assumption make any difference to the position of the image, I?

© Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014 **Photocopiable**

26.1 Investigating the current–voltage relationship in an ionic solution [M/M]; [O/R/R]; [A/I]

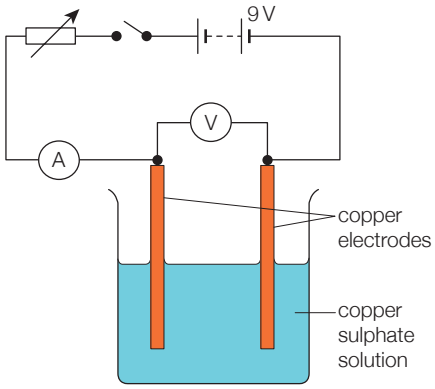
Chapter 26 More about electrical circuits

Syllabus reference D4

Introduction



Does an ionic solution, such as dilute copper sulphate solution, obey Ohm's law? You are to carry out an investigation into this preceding question. The apparatus shown in the diagram can be set up to investigate the current–voltage relationship for a dilute copper sulphate solution (an ionic solution) using copper electrodes.



Procedure

1 Connect up the circuit as shown in the diagram, but without making any connection to the power source.
2 Show your wired circuit to your teacher.
3 After your teacher has approved your wiring, connect the power source, under the teacher's supervision.
4 Vary the resistance of the rheostat so as to obtain a variety of current readings on the ammeter and corresponding voltage readings on the voltmeter.

Handling your results

1 Arrange your data in a suitable manner.
2 Plot a graph of current I (on the y-axis) against voltage V (on the x-axis) for the copper sulphate solution used in the experiment.
3 Extrapolate your graph and determine the minimum voltage required for electrical conduction to take place in the ionic solution.

Questions

1 Does the copper sulphate solution obey Ohm's law? Explain your answer.
2 Describe *two* kinds of experimental error involved in this experiment.
3 State *one* way in which electrical conduction in the copper sulphate solution differs from conduction in a copper wire.
4 On the same sheet on which you plotted your graph, sketch a graph expected if a more concentrated copper sulphate solution was used. Explain your reasoning.

Photocopiable © Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014

26.2 How much current could an LED safely handle? [P/D]; [M/M]; [O/R/R]; [A/I]

Chapter 26 More about electrical circuits

Syllabus reference D4

Introduction



You are given an LED (light-emitting diode) and are asked to find out by experiment the value of a suitable protective resistor, needed when using a 6 volt d.c. battery, that will allow the LED to light up fairly bright light without burning out the diode. You are NOT allowed to destroy the diode in this experiment.

Proposal

In chapter 30, in the section titled 'The p-n junction characteristic', a circuit for determining such a characteristic is shown in figure 30.18. A characteristic graph of diode current (*I*) vs. diode forward voltage (*V*) is shown in figure 30.19 (see page 476). By using a circuit similar to figure 30.18 (or other), plan and design your investigation. (Hint: If you can determine the maximum safe current from a diode current vs. diode voltage graph, you will then be able to calculate a suitable protective resistance.) Discuss your written plan with your classmates and then show it to your teacher.

Investigating and reporting

Once your teacher has approved of your plan, you are to go ahead with the practical activity, bearing in mind safety precautions and the need to not destroy the LED, while trying to obtain a bright light from it. In the Reflections section of your report, you may wish to comment on the suitability of the method for manufacturers to use in order to specify the protective resistance needed for that type of diode.

© Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014 **Photocopiable**

26.3 Series and parallel circuits [P/D]; [M/M]; [O/R/R]; [A/I]

Chapter 26 More about electrical circuits

Syllabus reference D4

Introduction



You need a 50 Ω resistor and a 200 Ω resistor as part of a project. However, all you have are 100 Ω resistors. The problem is that your circuit needs the resistances to be within ±10% accuracy (called 'tolerance'), but the resistances to which you have access have designated tolerances of ±20%. You wish to combine resistances to find out if your combinations will meet your tolerance requirement.

You are to design and carry out an investigation to find out if each 50 Ω and 200 Ω resistor combination you make will happen to meet the ±10% tolerance required by your project.

Proposal

You may refer to chapter 26 for a discussion on series and parallel circuits. In particular, figure 26.13 (in which a rheostat serves as the potential divider) and a voltage vs. current graph for the resistor might be useful. Upon designing your circuit (note that a graphical method is required) and the activities, discuss your plan with your classmates and then show your plan to your teacher.

Investigating and reporting

Once your teacher has approved of your plan, you are to go ahead with the practical activity, bearing in mind safety and accuracy precautions. In the Reflections section of your report, you may wish to comment on whether your combinations worked and whether the method can be counted on to work every time.

Photocopiable © Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014

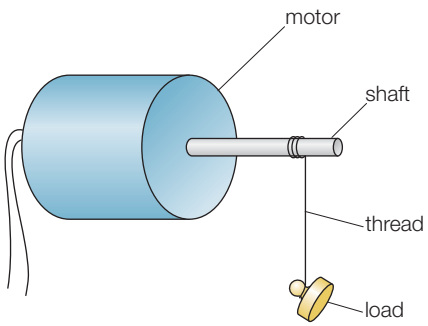
28.1 The efficiency of energy conversion in a d.c. motor [P/D]; [M/M]; [O/R/R]; [A/I]

Chapter 28 Electromagnetism

Syllabus reference D7

Introduction

A low-voltage d.c. motor can be used to lift a load that hangs form a thread attached to the motor shaft (see diagram). The work done in lifting the load is equal to the potential energy gained by the load: $PE = mgh$. The electrical energy supplied to the motor can be calculated using the equation: energy = IVt. You are to design an experiment to investigate the efficiency of energy conversion by the motor as a function of the load being lifted.



Proposal

1 Your experimental design should include the following:
- a statement of your hypothesis;
- a diagram showing how the motor will be supported;
- a circuit diagram showing how the motor will be connected to the power source and to the electrical measuring instruments that will be used;
- a list of equipment that you plan to use;
- a statement of variables to be controlled;
- a statement of the measurements to be taken, and how efficiency would be calculated from these measurements;
- a statement of any precautions you would take so as to obtain accurate results and also ensure safety.

2 Show your written plan to your teacher and then carry out the investigation.

Investigating and reporting

Once your teacher has approved of your plan, you are to go ahead with the practical activity, bearing in mind safety and accuracy precautions. In the Discussion or the Reflections sections, you may wish to comment on the following:

1 The extent to which your hypothesis was supported.

2 If the motor was not 100% efficient, what became of the energy that was not used for lifting the load?

3 What does the result of your investigation tell you about loads to be used on motorised equipment such as blenders?

4 What modification(s), if any, did you make to the original plan you submitted to your teacher? Give a reason or reasons for the modification(s).

© Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014 **Photocopiable**

35.1 Simulating radioactive decay [M/M]; [O/R/R]; [A/I]

Syllabus reference E3

Introduction

Radioactive decay is a random process. It can be simulated using a quantity of used matchsticks that are approximately square in cross-section and broken in half. (Alternatively, lead pencils with hexagonal cross-section can be used, either whole or cut up into small equal lengths.) One side of each half matchstick is marked (with a black dot, for example, as shown). The marked matchsticks, when placed in a container, represent 'undecayed nuclei'. However, when the matchsticks are poured on to a flat surface, those that land with the mark uppermost represent nuclei that have 'decayed'.

In this experiment, you are to determine the 'half-life' of a simulated radioactive decay process using marked matchsticks.

Procedure

1. Count a number of marked matchsticks (e.g. 500) and place them in a suitable container (e.g. a plastic bag, can or box). Record the initial count of 'undecayed' nuclei, and designate this throw as throw number zero (0).
2. After shaking the matchsticks in the container thoroughly, pour the matchsticks over a wide area onto a table-top. Remove all the 'decayed nuclei' (matchsticks that landed with their mark uppermost). Count and record the number of 'undecayed nuclei' present at the end of the throw, as well as the throw number.
3. Put the 'decayed nuclei' to one side. Replace the 'undecayed nuclei' in the container.
4. Repeat steps 2 and 3 for about 15 throws.

Handling your results

1. Plot a graph of number of 'undecayed nuclei' against 'time' (number of throws).
2. Using the graph in the previous step, determine the 'half-life' of the matchstick 'decay' process by considering:
   - **(i)** 500 as the initial number of 'undecayed nuclei';
   - **(ii)** 400 at the initial number of 'undecayed nuclei'.

Questions

1. Explain what is meant by the statement: 'Radioactivity is a random process.'
2. Why should the container with the matchsticks be shaken thoroughly before pouring out the matchsticks?
3. What was the 'half-life' of the process when:
   - **(i)** 500 was considered as the initial number of undecayed nuclei?
   - **(ii)** 400 was considered as the initial number of undecayed nuclei?
4. Does the half-life obtained in this experiment depend on the initial number of 'undecayed nuclei?' Explain your answer.

Photocopiable © Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014

35.2 Investigating radioactivity in food

Chapter 35 Half-life

Syllabus reference E3

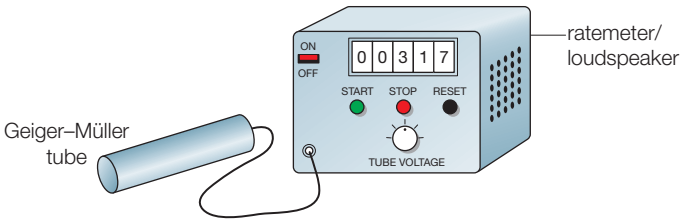
Introduction

If your laboratory is equipped with a portable Geiger–Müller tube and ratemeter or loudspeaker, you can use this equipment to investigate background radiation.

Note: You are *not* being asked to investigate known, highly radioactive sources, because of the dangers involved in doing so.

Background radiation and radiation from food are being investigated because they are ever present around us.

You are to plan and design an experiment to find out if a tea-bag emits more radiation than our normal surroundings.



Proposal

1 In your plan, indicate design aspects such as the following:
- a statement of your hypothesis;
- a statement of variables to be controlled and a description of how you will attempt to control these variables;
- a description of your method, including a clear indication of where and how readings will be taken;
- a diagram, or diagrams, of your proposed set-up;
- an indication of the number of readings to be taken.

2 Discuss your written plan with your classmates and then show it to your teacher.

Investigating and reporting

Once your teacher has approved of your plan, you are to go ahead with the practical activity, bearing in mind safety and accuracy precautions. In the Discussion or the Reflections sections, you may wish to comment on the following:

1 Was there any noticeable difference between the amount of radiation from the tea-bags and that from the background? Did the results you obtained support your hypothesis?

2. What if your results indicated that the tea-bag emitted more radiation than the background? What should you do?

3 Is it important, in experiments of this nature, to take counts over a very large, as compared to a small, period of time? Explain your answer.

4 What modification(s), if any, did you make to the original plan you submitted to your teacher? Give a reason or reasons for the modification(s).

5 If time permits, you may want to test other foodstuffs such as bananas or nuts.

© Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014 **Photocopiable**

Appendix

The Greek alphabet

lower case	upper case	name	pronounced
α	Α	alpha	alpha
β	Β	beta	bee-ta
λ	Γ	gamma	gamma
δ	Δ	delta	delta
ε	Ε	epsilon	ep-si-lon
ζ	Ζ	zeta	zee-ta
η	Η	eta	ee-ta
θ	Θ	theta	thee-ta
ι	Ι	iota	eye-oh-ta
κ	Κ	kappa	kappa
λ	Λ	lambda	lam-da
μ	Μ	mu	mew
ν	Ν	nu	new
ξ	Ξ	xi	k-sigh
ο	Ο	omicron	oh-mi-cron
π	Π	pi	pie
ρ	Ρ	rho	roh
σ	Σ	sigma	sigma
τ	Τ	tau	taw
υ	Υ	upsilon	up-si-lon
φ	Φ	phi	fi
χ	Χ	chi	ki (like sky)
ψ	Ψ	psi	p-sigh
ω	Ω	omega	oh-mega

Photocopiable © Alec Farley and Clarence Trotz 2014. Design and illustration © Macmillan Publishers Limited 2014

Index

90°–45°–45° prism 331–2

absolute zero 187–8, 231, 233
absorbers 218–19
acceleration 6
  definition 105
  and force 113, 121
  formula 109
  and power 137
  ticker-timer measurement 118
  uniform 107
accepted (expected) value 10, 57
acidic gases 539
action and reaction 114
action of points 378
adhesion 171
air
  conduction in 216
  convection in 217
air-conditioning 209, 222, 223
air traffic control 473
alarm circuit 478, 484, 486, 491
alcohol thermometer 191, 192–3
alkaline primary cells 461, 464
alpha decay 524
alpha particles 499, 500, 519, 521–3
  uses of 532–3
alternating current (a.c.) 387
  in an oscilloscope 472, 477
  for energy transfer 446
alternating current (a.c.) generator 440–1
alternative energy sources *see* renewable energy sources
alternator 440–1
Amaila Falls Hydro Project 142
americium 529, 532–3
ammeters 17, 391–2
  errors in 29–30
  in a series circuit 400
  using 407–8
ampere 4, 386
amplifier 266
amplitude 54, 256, 273–5
  and loudness 279
AN rule 422, 437
analogue instruments 15, 33
  zeroing 30
analysis and interpretation skills (A/I) 47, 564–80, 580
AND gate 487–8, 489
angle
  of diffraction 277
  of reflection 41, 270, 308–9
  between vectors 67
angle of incidence 41
  and angle of refraction 329
  in reflection 308–9
  in refraction 319, 320, 323, 325
  in sound waves 270–1
  in total internal reflection 331
angle of refraction 271, 319, 320, 327, 328, 344–5
  and angle of incidence 329
  colours 333
angular displacement 54–5, 60
annular eclipse 302
anode 461, 469–70, 475–6
anomalous expansion 179–80
anti-parallel forces 87–9
  lever principle 91, 92
anti-parallel vectors 70–1
aperture size 350–2
apparatus 556
  handling and manipulating 31
  reporting 34
  sensitivity 33
  see also instruments
archaeological dating 534–5
Archimedean upthrust 81
Archimedes' principle 156–7
  and relative density 158–9
  sinking and floating 159–62
area 6
  of contact 147–9
  under a speed–time graph 107
  under a velocity–time graph 107, 109–10
Aristotle 53, 104
arm of a moment 90
armature 435
atmospheric pressure 177–8
atomic models 374
  Bohr's 501–3
  Dalton's 497
  early ideas 496–7
  Rutherford's 499–501
  shell 502, 510–11
  Thomson's 497–8, 498, 499
  wave 504
atomic number 508, 509, 512–13
atomic spectra 503
atoms 168, 373, 496
  bonding between 511–12
  stability of 507–9
audible range 279
audio frequency oscillator 275–6
Avogadro number 5

background count rate 522, 529–30
bagasse 142
baggage inspection 534
balance (instrument) 16, 30
ball-and-ring experiment 178
bar magnet 418, 419
base units 3, 4–5
batteries 381, 385–6, 388, 396
  primary and secondary 460–3
beam balance
  equilibrium 89–90
  lever principle 91–2
Becquerel, Henri 516, 517
becquerel (Bq) 520
beta decay 524
beta particles 519, 521–3
bimetallic strips 179, 190
binary system 487
binoculars 331, 332
biomass 141, 142
'bits' 15–16
black surfaces 218, 219
block-and-tackle pulley system 564–5
blood pressure 155
bob 54–6
  displacement of 59–60
  mass of 61–2
body temperature 209
Bohr's atomic model 501–3
boiling 175, 177–8
boiling point 174
  factors affecting 177–8
Bourdon gauge 227, 228
Boyle's law 227–30
brass 196
breeder reactors 545
breezes 217, 223
Brownian motion 171
buoyancy 156
  positive and negative 159, 161
burglar alarm 484, 486

C-shaped magnet 420
caesium clock 4
calibration 186
caloric theory of heat 196
calorie 196, 197
calorimeter 200, 202
cameras
  digital 297–8
  pinhole 303–5
cancer treatment 521, 534
candle wax 174–5
cannons 196
capacitance 381
capacitors 380–1
car engine 226, 227
car starter circuit 424, 485
car tyre pressure 234
carbon-12 513, 534
carbon-14 513, 524, 529
  dating 534–5
carbon brushes 434–5, 440, 441
carbon dioxide 220
Caribbean
  building ventilation 222–3
  hurricanes 132, 133, 142
  power generation 450, 451, 464
cathode 461, 469, 475–6
cathode-ray oscilloscope 275, 469–70
  some uses 471–3
  using 470–1
cathode-ray tube 469–70, 497, 498
cathode rays 469, 497–8
cells 388
  comparing 464
  making 396
  positive and negative terminals 390
  primary and secondary 460–3, 464
Celsius scale 186, 187, 231
  converting 188
centigrade scale 187
centre of gravity 80–1, 88, 96
centre of mass 81, 88
  finding position of 95–7
  lamina 96–7
  and stability 100–1
centre-zero galvanometer 437–8, 442
Chadwick's experiment 500–1
chain reaction 541
change of state 174–80, 204
charcoal 141
charge-to-mass ratio 507
charging
  capacitor 381
  by contact 375
  electron theory of 374
  by friction 375–6
  by induction 376–7
Charles' law 231–2
chemical potential energy 131, 133, 134, 135
Chernobyl, Russia 544, 546
chief ray 347, 348–9
  in ray diagrams 355
chlorofluorcarbons (CFCs) 208
circuit-breaker 455, 456, 458
circuit diagram 389, 391, 392, 395, 557
circuit symbols 393

circuits *see* electric circuits
circular waves 258
claw-hammer 99
cleavage planes 169–70
clinical thermometer 191–2
clocks 4, 53
cloud chamber 518–19
coal 539
cobalt-60 529, 533
cognitive skills 30, 47
coherent sources 273, 275
cohesion 171
coil orientation 440–1
colour-coding of wires 454
colours 289
  dispersion 333–5
  emission of 503
  Newton and 335–6
  of objects 337–8
  recombination 336–7
commutator 435, 441
components 389
compressibility 173, 174
compression (soundwaves) 264–5
compression (waves) 244, 249, 254, 255
concave lens *see* diverging lens
condensation 174, 175
condenser 450–1
conduction 213, 214–16
  applications of 221
  electron theory of 374–5
  refrigeration and 209
conductors 215, 220, 374
  distribution of charge on 378–9
  good and poor 389–90
  lightning 379
congruent triangles 314
constant-volume gas thermometer 233
constructive interference 274, 275–6, 292, 299
consumer unit 455–6
contact forces 76, 77
continuous spectrum 285
contraction 178–9
control rods 544
controlling variables 44
convection 213–14, 216–17
  applications of 221, 223
  refrigeration and 209
convection current 216–17, 223
conventional current 386
convergent beams 300
converging lens 343, 344–8
  aperture size and 350–2
  focal length 352–3, 571–2
  focus 344
  images from 359–61
  magnification 362
  ray diagram construction 355–8
  thickness of 352
  virtual image location 354–5
converging rays 246, 313
cooking 220–1
coolant 544
cooling
  curves 46
  evaporation and 176
  theory of 44
coplanar forces 87–8, 92
coplanar vectors 66, 67
corrected count rate 522, 530
coulomb (C) 373, 387
count rate 520, 522, 529–30
counter 519, 520
covalent bonds 512
crests 242, 243, 245, 257
critical angle 329–30, 331
critical mass 543
crow-bar 99
crystalline structures 170
CS rule 422, 438
Curie, Marie and Pierre 517
curie (Ci) 520
current *see* electric current
current-carrying conductors
  current–voltage characteristics 409–10
  electromagnetic force 431–3
  magnetic field around 421–3
current electricity 371, 385
current–voltage characteristic 409–10, 476
current–voltage relationships 407–8
  determining 408–11, 576
  see also electric current; potential difference
cycle of a wave 242, 253, 254–5
cycles per second 253

Dalton's atomic model 497
decay curve data 530, 531, 534, 535
decay equations 524
decay rate 529
decommissioning 546
degree of precision 11–13
degree of significance 12, 14, 19–20, 28
  using a formula 20
degree of temperature 187
demagnetising 427
density 6, 7, 179–80
  and floating 161
  and hydrostatic pressure 149–51
  relative 7–8, 158–9, 161
  and sinking 159, 160
  ultrasound 280
dependent variable 28, 29, 33, 34, 46
depolariser 461
depth, and hydrostatic pressure 149–52, 155–6
depth-sounding 280
derived quantities 6–7
destructive interference 274, 276, 292, 299
deuterium 542
deviation 324, 325, 332, 334, 344–5
diagram, labelled 34, 44–5
differences, and limits of error 18–19
diffraction 276, 291–2, 299
  at an edge 278
  at an opening/gap 277
diffusion 170–1
digital camera 297–8
digital instruments 15–16, 33
digital multimeter 408
digital on–off switching 492
diodes
  p-n junction 475–6, 477
  photodiode 298
  properties 485–6
  see also light-emitting diode (LED)
dipoles 418, 425
direct current (d.c.) 387
direct current (d.c.) circuit 389, 395–6
  potential difference from 402–3
  using ammeters 407
  using oscilloscope 477
direct current (d.c.) generator 441
direct current (d.c.) motor 434–5, 579
directly out of phase 255, 256
directly proportional 228, 231, 232
discharging 376
dispersion 333–5
displacement 65–7, 256
  definition 105
  out of phase 274, 275
  in phase 273–4
  positive and negative 111
  in a velocity–time graph 107, 109–10
displacement–position graph 248–9
displacement–time graph 106–7, 110–11
  sound waves 274
  waves 246–8
distance covered 67, 105, 564–5
distance–time graphs 105–6, 110–11
distribution box 455–6
divergent beams 300, 305
diverging lens 343, 344–5, 347–8
  focal length 354–5
  focus 345
  images from 359–61
  magnification 362
  ray diagram construction 358
diverging rays 246, 313, 314
domain theory of magnetism 425–6
domestic hot-water systems 217
doping 474–5
dosimeters 521
dots and crosses convention 422
double-glazing 216
drag 82
dust precipitator 378
dynamic equilibrium 176
dynamics 111
  and Newton's laws of motion 112–17
dynamo 441, 451

Earth
  gravitational field strength 80, 91
  gravitational force 80–1
  as a magnet 421
earth wire 454–5
earthing 376, 377
echoes 268–9, 270
eddy currents 439
effective resistance 406
efficiency 138–40, 564
  energy conversion 579
effort 97–8, 99, 138
Einstein's quantum theory of light 297
Einstein's theory of relativity 540
elastic behaviour 172
elastic collisions 169
elastic force 76, 77, 83–4
elastic potential energy 131, 134
electric bell 425, 426
electric cars 142
electric charge 6, 371
  calculating 387
  detecting and estimating 377
  distribution on conductor surface 378–9
  forces between 373
  negative and positive 372–3
  storing 380–1
  see also charging
electric charge carriers 386
electric circuits
  describing 390–5
  energy and 388–90
  energy and power in 412
  household 453–9
  low frequency e.m. production 288
  see also direct current (d.c.) circuit; parallel circuits; series circuits
electric current 386, 391–2
  base units and symbols 3
  definition 387
  hazards of 385
  household equipment 457
  LED handling 577
  magnetic field around 421–3
  in a motor 434–5
  overhead power cable 411
  in parallel circuits 403–5
  in series circuits 400
  in series–parallel circuits 406–7
  from a transformer 444

see also alternating current (a.c.); current–voltage relationships; direct current (d.c.)
electric field diagrams 380
electric fields 285, 379
direction of 380
radiation emissions 522, 523
electric shock 385, 454
electric stove 412, 458
electrical energy 285, 450
conserving 464
electrical hazard warning 385, 457
electrical methods
magnet-making 426
specific heat capacity of liquids 200–2, 567
specific heat capacity of metals 202–3
specific latent heat of fusion 206
specific latent heat of vaporisation 207–8
electrical outlets 456
electrical wiring in a house 454–9
electricity 371
distributing 452–3
generating 450–2
good and poor conductors 389–90
and magnetism 417, 431
magnetism from 421–5
paying for 459–60
portable 460–4
safety and 385
at work 450
electricity meter 455–6, 460
electrodes 461
see also anode; cathode
electrolysis 389
electrolytes 388
electromagnetic energy 285, 297
electromagnetic force 83, 431–4
applications 434–6
size of 434
electromagnetic induction 431, 436–9
applications 440–6
Faraday's experiments 436
in a metal sheet 438–9
mutual 442
in a straight wire 439
electromagnetic radiation 218, 284
absorption of 219
frequency 287, 288, 297
general properties 291–2
types of waves 285–6
electromagnetic relay switch 485
electromagnetic spectrum 286–8
electromagnetic waves 243, 250
nature of 284–5
production and uses 284
properties and uses 288–91
solar 285, 286
electromagnetism 431
electromagnets 424–5, 451
electromotive force (e.m.f.) 390, 391
induced 436–7
lead-acid cell 462
Leclanché cell 461
electron current 386
electron theory 374–5
electronic circuits
diode properties 485–6
p-n junction diode 476
semiconductor resistors 486
electronic switches 483, 485, 487, 491–2
electronic thermometers 189–90, 192, 478
electronics 468, 483
damage to 459
simple applications 485–91
electrons 373, 468, 501
charge on 506–7, 509
discovery of 498
flow of 388
free 214, 215, 374, 473–5
mass of 507
orbitals 504
orbits 374
shells 502–3, 510–11
stable energy states 508
valence 473–5
electrostatic
attraction and repulsion 372
dust precipitator 378
photocopier 372–3
spray painting 373
electrostatic force 76, 82, 508–9
electrovalent bonds 512
elementary charge 506
elements 497, 508
Periodic Table 509–10
emergent ray 322, 323, 324, 345
emission spectrum 503
emitters 218, 219
endoscopy 333
energy 6
in an electric circuit 388–90, 412
and efficiency 138–40, 564
and friction 134–5
law of conservation of 133–4, 139
meaning of 127
from nuclear fusion 542
sources of 141–2
see also various types of energy
energy conversion 389, 579
energy equivalent 545
energy flow diagram 133, 134
energy levels 502–3, 508
energy transfer
alternating current for 446
theory 197, 198
entrance unit 455, 456
equilibrium 88–9
of a beam balance 89–90
dynamic 176
errors
limits of *see* limits of error
parallax, random and zero 18, 27–8, 29–30
percentage 267, 268–9
reaction 17, 21, 61, 267, 268–9
shown in graphs 29
spreading 21, 269
systematic 29–30
see also uncertainty
evaporation 175, 178
cooling effect of 176
factors affecting 176–7
expansion 179–80
expansion valve 209
extended source of light 301, 302
extrinsic semiconductors 475

Fahrenheit scale 186, 187
converting to 188
fairy lights 402, 405, 453
farad (F) 381
Faraday, Michael 431, 436, 442
Faraday's law of electromagnetic induction 437
fast medium 326
Fermi, Enrico 541
ferromagnetic materials 425
field-effect transistor (FET) 485, 491–2
field lines 380
field strength 380
fire alarm 179, 478, 486
fire extinguisher 385
fixed capacitors 380
flashlight 389
Fleming's left-hand (motor) rule 432–3, 434, 435, 523
Fleming's right-hand (generator) rule 439, 440–1
floating 160–1
flux density 433
fly-back 471
focal length 348, 350
calculations 365–6
converging lens 352–3, 571–2
diverging lens 354–5
from scale construction 364
focal plane 348–9
focal point 347
forces 6
and acceleration 113, 121
of attraction and repulsion 173
between charged particles 373
definition 76
line of action 77
and liquid pressure 153–4
between particles 171–2
and power 137–8
production 75–6
at rest 88
types of 127
units of 7
as vector quantity 68–71
and work 128–9, 130, 564
see also various types of forces
forward biasing 476
forward voltage 410
fossil fuels 220, 450–1, 464
future concerns 539
global dependence on 538
Foucault, Jean 296
fracking 539
free electrons 214, 215, 374, 473–5
freezing 174, 175
frequency 54, 253, 254
audible range 279
electromagnetic radiation 287, 288, 297
measuring 472
pitch and 279
sensitivity to 276
solar waves 286
ultrasound 280
in the visible spectrum 333
friction 82, 99
charging by 375–6
and energy 134–5
energy lost through 139
and heat 196, 197
fringes 298–9
fuel rods 543
Fukushima, Japan 544, 546
fulcrum 89, 97, 98
fundamental quantities 3
derived 6–7
furnace 450
fuse rating 394–5
fuses 394, 455, 456, 458

Galileo Galilei 53, 105
gamma decay 524
gamma rays 285, 286, 287, 288, 292
nature and properties 521–2
range 522–3
tracks of 519
uses of 533–4
gas equation 234–5
gas pressure 155, 172
gas station fires 376
gases 174, 226
Boyle's law 227–30
change of state 175–6
characteristics of 174
Charles' law 231–2
pressure law 232–4
gasohol 142
Geiger–Marsden experiment 499
Geiger–Müller tube 519–20, 522, 529
generating station 445, 453
generators 440–1, 545
geocentric theory 53
glancing angles 42, 309
glass block 319–20, 321–3, 327, 334
global warming 220

graduated (calibrated) scales 15, 17
grandfather clock 53
graphs 560
  assessment criteria 40
  design of 33
  line of best fit 28–9, 34
  plotting 34, 35, 38
  straight line graph theory 57
  see also different types of graph
gravitational field 80
gravitational field strength 80, 81, 91
gravitational force 76, 77, 78–9
  Earth 80–1
gravitational potential energy 129–31, 133, 134, 135, 136
gray (Gy) 521
greenhouse effect 219–20
greenhouse gases 539
grid 470
groups 509, 510–11

half-life 525, 528–9
  constancy of 531–2
  practical determination 529–31
hammer 138
head of liquid 155
heart pacemakers 528, 533
heat
  caloric theory of 196
  Count Rumford's experiments 196–7
  Joule's experiment 197
  latent 175, 204
  see also thermal energy
heat capacity 199
  see also specific heat capacity
heat energy 185, 198, 450
  and specific heat capacity 198–9
heat exchanger 222, 544
heat pumps 223
heavy water 544
hectare 5
heliocentric theory 53
helium atom 508
hertz (Hz) 253
high-intensity filament lamps 410
high-power circuits 456, 458
'holes' 386, 468, 473–5
Hooke's law 88–9
horseshoe magnet 420
hotness 184, 186
household electric circuits 453–9
household energy requirements 540
household equipment, electrical requirements 457
hurricanes 132, 133, 142
Huygens' wave theory of light 296, 299
hydraulics 153
hydrogen atom 508
hydropower plants 142, 452
hydrostatic pressure 149–62
hypothesis 43–4, 59, 186

ice 180, 204–6, 210
ice point 187
images
  distance 351–2, 363, 366
  formed by glass prism 573–5
  formed by plane mirrors 309–15
  pinhole camera 303–4
  quality 350–2
  size 304, 313–14
  see also real image; virtual image
implementation 558–62
in antiphase 255, 256
in phase 254–5, 256
  electromagnetic wave 285
  sound wave 273–4
incident rays
  deviation 345
  in reflection 308–9, 310, 314
  in refraction 319, 320, 322, 323, 324
independent variable 28, 29, 33, 34, 46
index notation 8
induced current 436, 437–8
induced e.m.f. 436–7
induced magnetism 426
induction
  charging by 376–7
  see also electromagnetic induction
inertia 78, 138
inference 35, 78
infrared (IR) radiation 218, 219–20, 285, 287, 289–90
infrared thermometers 193
input voltage 443
  controlling 487–8
instruments
  analogue 15, 30, 33
  digital 15–16, 33
  precision 11–13
  spreading the error 21
  uncertainty of 10–11
  uniform scale 13–14, 17
  zeroing 30
insulation properties 42–7
insulators 216, 374
  jacket 201
integrated circuit 479, 487
inter-conversion of energy 133–4, 135
interference 291–2
  of sound waves 273–6
  Young's slits experiment 299
internal combustion engine 226, 227, 451
internal energy 184–5, 198
intrinsic semiconductors 474
inverse relationship 148–9, 227, 228
inverted image 358, 359–61
inverter gate 488–9
iodine-131 529, 533
ionic bonds 512
ionic lattice 511
ionic solutions 410, 576
ionising power 522, 525
ionosphere 291, 292
ions 168, 169, 214, 215, 386
  positive and negative 378
iron-53 535
iron filings 417–18, 419
isotopes 513
  radioisotopes 528, 532–5
Iter project 546

Joule heating 439
joule (J) 90, 127, 128, 130, 199
joulemeter 200
Joule's experiment 197

kelvin (K) 4, 188, 199, 231
Kelvin scale 186, 187–8, 231
kilocalorie (kcal) 196
kilogram metre per second 113
kilogram standard 3
kilogram(kg) 3
kilowatt-hours (kWh) 460
kinematics 53, 111
kinetic energy 127, 132–3
  and change of state 175, 176
  and heat 198
  inter-conversion 133, 135
  in vibrating systems 184
kinetic theory of matter 169, 198
  applying 172–4
  and Boyle's law 230
  and Charles' law 232
  evidence for 169–72
  and the pressure law 234
knowledge and comprehension (KC) 31

laboratory thermometers 191
lamina 96–7
lamps 389, 458
  colour of light from 503
  connected in parallel 454
  current passing through 392
  energy converted by 391
  high-intensity filament 410
  lower energy 464
latent heat 175, 204
  see also specific latent heat
lateral displacement 323, 324, 346, 347
lateral inversion 314–15
law of charges 371
law of energy conservation 133–4, 139
laws of reflection 269, 308–9
  first 270, 308
  second 270, 308
  verification 41–2
laws of refraction 320–1
lead-acid cell 461–2, 464
Leclanché dry cell 460–1, 464
length
  base units and symbols 3
  pendulum 54–9
lens equation 365–7
lens stop 353
lenses
  action on parallel rays 344–5
  definition 343
  pure spectrum from 336
  scale construction 362–5
  shapes of 343
  spherical 344
  terminology 346–9
  see also converging lens; diverging lens
Lenz's law 437–8
Leslie's cube 219
lever principle 90–3
  applied in machines 97–8
levers, classes of 98, 100
light
  rival theories of 296
  speed of 267, 325, 326
  wave behaviour 298–9
  wave–particle duality 296, 297
  see also white light
light-dependent resistor (LDR) 478–9, 486, 488
light-emitting diode (LED) 388, 477
  logic gates 487–8, 489
  polarity 395
  protecting 393, 577
light-meter 478–9
light rays 246, 333
  refraction at an interface 327–9
  through a block 319–20, 321–3
  through a prism 323–4
  travelling in straight lines 300–1
  see also parallel rays
light spectrum *see* visible spectrum
lighting circuits 456, 458, 486, 489
lightning 267, 379, 459
lightning conductor 379
limits of error 13, 16, 270
  sums and differences 18–19
limits of uncertainty 16
line of action 77
line of best fit 28–9, 34
linear relationship 190
linear scale *see* uniform scale
liquid-in-glass thermometer 179, 190–1, 192
liquids 173–4
  boiling 175, 177–8
  change of state 174–5
  conduction 214–15
  connected to each other 152
  diffusion 170–1
  evaporation 175–6
  hydrostatic pressure 149–62

in open vessels 149–52
specific heat capacity 200–2, 203–4, 209, 567
unconfined 153–4
lithium cells 461, 462, 464
live wire 454, 455, 458
load 97–8, 99, 138, 565
lodestone 417
logic circuits 487
logic gates 487–91
logic state 487–8
longitudinal waves 243–4, 249
displacement 274
period 254
sound 265
loudness 279
loudspeaker 266, 276
low voltages 459
lower fixed point 187, 189–90
lunar eclipse 303

machines, lever principle 97–8
magnetic energy 285
magnetic field strength 433
magnetic fields 285, 418–21
around a current-carrying wire 421–3
electromagnetic force 431–3
mapping 420
radial 435–6
radiation emissions 522, 523
uniform and non-uniform 420
magnetic flux 419, 433–4
and induced e.m.f. 437
magnetic flux linkage 437
magnetic force 76, 83, 418
magnetic lines of force 419, 420, 423
magnetic materials 417
magnetic poles 417, 418, 421, 422, 424, 438
identity of 419
magnetic switches 484
magnetically hard/soft materials 425
magnetism 417–18
domain theory 425–6
and electricity 417, 431
from electricity 421–5
magnetite 417
magnets 417–18
electromagnets 424–5
magnetic fields around 418–21
methods of making 426–7
permanent and temporary 425–6
magnification, image 304, 359–61, 362, 363
main fuse 455
main switch 455–6
manometer 154–6
mass
base units and symbols 3
centre of *see* centre of mass
definition 78
and gravitational force 80
and kinetic energy 132–3
scalar 64
mass defect 540, 541
mass-energy 540
mass number 512–13
matter waves 243, 249–50
maximum-minimum thermometer 193
mean (average) value 22–3, 28, 30
measurement, units of 2
measurement and manipulation (M and M) skills 40–2, 564, 569–80
mechanical advantage 99
mechanical energy 250, 266
mechanical equivalent of heat 197
mechanical forces 76, 77, 81–2
mechanical switches 484
medical imaging/diagnosis 280, 288, 290, 332, 333
meltdown 546
melting 175
melting point 174
mercury oxide cells 461, 464
mercury thermometer 191
metal-leaf electroscope 377, 378
metal oxide varistor (MOV) protectors 459
metals 510
bonding 511
conduction 214, 215, 374
specific heat capacity 202–4, 568
thermal conductivity 215–16, 220
method of mixtures 203–4, 205–6, 566, 568
metre 4
metre rule 12, 13, 21
metre scale reading 316
micrometer screw gauge 12, 14, 30
micron 5
microphone 266, 275–6
microwave cooking 220–1, 290–1
microwaves 285, 286, 287, 288, 290–1
diffraction and interference 291
milk 567
milliammeter 189, 291, 478
milligram (mg) 7
millimetre (mm) 4
mind mapping 42–3
mirror *see* plane mirror
mirror image relationship 315
mirror line 314
moderator 544
molecular theory 172–3
molecules 168, 169
conduction 214–15
in gases 230, 232, 234
moment
arm of a 90
of a force 89
law of 93
and stability 100
momentum 6, 112
in collision 115–17
conservation of 113, 120
magnitude of 172
Moon, gravitational force 81
motion in a straight line 105
moving-coil galvanometer 420, 435–6
multimeter 189, 408
multiple units 4–5
mutual induction 442

N-pole 418, 419, 421, 422, 424, 438
n-type semiconductor 475
NAND gate 489–90
national grid system 453
negative charge 372–3
negative terminal 388, 390, 395
neutral point 419, 433
neutral wire 454
neutron capture 541, 542
neutron number 512
neutrons 373, 500–1, 508, 541
charge on 507
mass of 507
newsprint, insulation and 42–7
Newton, Isaac 112
and colour 335–6
Opticks 296, 335
Principia Mathematica 53, 112
scientific methodology 553
Newton disc 337
newton-meter 17, 33, 90, 94–5, 564
newton (N) 6, 7, 80, 117
newton per kilogram 80
Newtonian telescope 335
Newton's laws of motion
first law 112–16, 119–20
second law 113–17, 120–1, 159, 172
testing by experiment 117–21
third law 114–17
Newton's particle theory of light 296
nickel-cadmium (nicad) cells 462, 464
no-parallax 311–12, 354
non-contact forces 76, 77, 82–3
non-linear scale 17
non-metals 510
non-parallel forces 87
NOR gate 490
normal 271, 272, 308–9, 319, 320
normal reaction 81
normally closed/open switch 484
NOT gate 488–9, 490
nuclear bomb 541, 546
nuclear energy 539–47
accidents 544
nuclear fission 541, 542
nuclear fission power plant 543–5
nuclear force 76, 84
nuclear fusion 542
nuclear fusion reactors 545
nuclear 'planetary' atomic model 499–501
nuclear power debate 545–6
nuclear radiation 545–6
nuclear reactor 543–4
nucleon number 512
nucleons 507
nucleus 373, 499, 512–13
radioactive emissions from 521–2
stability of 508, 524
structure of 500
nuclides 512–13, 524

object distance 312, 351–2, 362, 365, 366
observing, recording and reporting (ORR) skills 36–40, 554–5, 564–80
Oersted's discovery 421, 431
ohm 392
Ohm's law 392
optical centre 346, 347, 352, 356
optical density 326
optical fibres 332–3
Opticks (Newton) 296, 335
OR gate 488, 490, 491
orbitals 504
ordinary inversion 315
orientation, image 304, 314
oscillations 241, 246–8
definition 54
electromagnetic 284–5
and frequency 253
out of phase 274, 275
pendulum 22–3, 53, 56–9
and period 253
in phase 273–4
oscilloscope *see* cathode-ray oscilloscope
out of phase 255, 256, 274, 275
outermost shell 510
output voltage 443
overhead power cable 411
overloaded circuit 394–5, 459

p-n junction characteristic 476
p-n junction diode 475–6
rectification using 477
p-type semiconductor 475
paired forces 77, 80
paper sheet manufacture 533
parallax 312
parallax errors 18, 27–8, 316
parallel beams/rays 300
parallel circuits 578
current and voltages in 403–5
household 453–4
resistance in 405
with series circuits 406–7
parallel connection 390

parallel forces 87
  equilibrium at 88
  resultant of 94–5
parallel rays
  distant objects and 349–53
  lenses action on 344–5
  see also light rays
parallel vectors 70–2
parallelogram method 67–70
paraxial rays 348–9, 353
  in ray diagrams 355
partial eclipse 302, 303
partial reflection 330, 331
particle models 173
pascal (Pa) 7, 146–7, 148
Pascal's law 153, 155
path difference 257, 276
peak voltage 471
pendulum 241
  displacement of bob and period 59–60
  experiment 35, 47
  length of pendulum on period 54–9
  mass of bob and period 61–2
  period of oscillation 22–3, 53
  simple 53–4
penumbra 301, 302, 303
percentage error 267, 268–9
period 253–4, 472
  and bob displacement 59–60
  and bob mass 61–2
  oscillations 241, 246
  and pendulum length 54–9
Periodic Table 509–11
periods (elements) 509, 510, 511
periscopes 331–2
permanent magnets 425–6
Perspex 569–70
phase *see* in phase; out of phase
phase difference 257, 276
phosphor 470
phosphorus-32 529, 533
photocopiers 372–3
photodiode 298
photoelectric effect 297
photons 288, 296
  in modern technology 297–8
photovoltaic cells 463
physical quantities 2–3
pinhole camera 303–5
pistons 154, 226, 227
pitch 279
pivot
  centre of mass 95–7
  reaction at 92–3
pixels 298
place value 12–13
plane mirror
  image formed by 309–15
  uses of 316
  virtual image location 354
plane of incidence 308, 314
plane rays 246
plane waves 258, 276–7
planning and designing (P and D) skills 42–7, 553–4, 566–8, 578–9
plastic, insulation properties 42–7
Plimsoll line 160
plotting compass 418, 419
plugs 454
plutonium 545, 546
point charges 380
point of incidence 308, 319, 320
point source of light 300–1
polarisation 461
pollution 539
polonium 517, 529
portable electricity 460–4
positive charge 372–3
positive terminal 388, 390, 395
potassium manganate (VII) 170–1, 217
potential difference 6, 390, 391
  from a d.c. source 402–3
  household wiring 457
  measuring 471–2
  overhead power cable 411
  in parallel circuits 403–5
  in series circuits 401
  see also current–voltage relationships; voltage
potential dividers 393, 402–3, 409, 487
potential energy 127
  inter-conversion 133–4, 135
  types of 129–31
  in vibrating systems 184
  work done 136
power 6, 135–8
  in an electric circuit 412
  formula 136
  from a generating station 445, 453
  input and output 139
  from a transformer 444
  unit of 137
power losses, minimising 452, 464
power ratings 457
power station generators 441
precision 11–13
  digital instruments 16
  spreading the error 21, 269
  uniform scale instruments 13–14, 17
pressure 6, 146
  applied to liquids 153–5
  definition 146
  gas 155, 172
  hydrostatic 149–62
  solids and 147–9
  standard 187
  units of 7, 146–7
  vs temperature of gas 232–4
  vs volume of gas 227–30
pressure cooker 177
pressure law 232–4
pressure vessel 544
primary cells and batteries 460–1, 463, 464
primary coil 442, 443–4
primary colours 337
primary source of energy 134, 141
primary standards 3
principal axis 344–5, 346, 347–8, 356–7
principal focus 344–5, 347–8, 356
principal plane 346, 347, 356
Principia Mathematica (Newton) 53, 112
principle of flotation 161
principle of superposition 273, 275
prisms 323–4, 331–2
  deviation of rays 344–5
  image formed by 573–5
  white light dispersion 334, 335
product/sum formula 405, 406, 408
proof plane 378
proposal 556–8
proton number 508, 512
protons 373, 500, 501, 508
  charge on 507, 509
  mass of 507
psi (pounds per square inch) 234
psychomotor skills 30
pull 75–7, 80, 82
pulse 241–4
pure number 7
pure spectrum 336
push 75–6, 77, 81
push-back 77
Pythagoras' theorem 66

quantum 297, 502
quantum numbers 502, 504

radar 291, 473
radial magnetic field 435–6
radiation 214, 217–18
  applications of 221
  emission of 219
  refrigeration and 209
  see also electromagnetic radiation; infrared (IR) radiation
radiation sickness 517, 521
radiation warning sign 517
radio waves 285, 286, 287, 288, 291
  diffraction and interference 291–2
radioactive decay 523–4, 529
  simulating 531–2, 580
radioactive emissions
  detecting 518–20
  making use of 525
  types of 521–3
radioactive materials, safety and 517–18
radioactive tracers 533
radioactivity 516–17
  in food 581
  random nature 520
  units of 520–1
radioisotopes 528
  uses of 532–5
radiotherapy 521, 534
radium 517, 524
radon 524, 529, 530–1
rafts 161
rainbow 334
random errors 18, 27–8
randomly occurring events 520
range, radiation emissions 521, 522–3
rarefaction 244, 249, 254, 255
  sound waves 264–5
ratemeter 519, 520, 530
ray diagrams 304, 305, 313
  constructing 355–8
  lenses 347
  in scale construction 362–3
ray optics 300, 315
ray tracing 310–11
rays 245–6
  and wavefronts 258
  see also different types of rays
reaction error 17, 21, 61, 267, 268–9
read–write head 420, 426
real focus 344, 347
real image 304, 313, 359–61
recharging 462, 463, 464
rectangle method 72–3
rectification 477
rectilinear propagation 300
reed relay switch 485
reference point 56
reflected ray 308–9, 310, 313, 314
reflection
  of sound 269–70
  total internal 330–1
reflectors 218
refracted ray 319, 320
refracting surface (boundary) 319
refraction of light 319
  dispersion into colours 333–5
  general formula at interface 327–9
  in transparent media 325–6
refraction of sound 270–2
refractive index 325–6
  and critical angle 329–30
  different media 328
  glass block experiment 327
  Perspex 569–70
refrigerant 208, 223
refrigerators 208–9
relative atomic mass 509
relative density 7–8, 158–9, 161
relay switch 485
relay switching 424–5, 426
renewable energy sources 141–2, 451–2, 464, 539
reservoirs 452

resistance 392–3
and fuse rating 394–5
in parallel circuits 405
and power losses 452
in series circuits 401–3
in series–parallel circuits 406
of a wire 556–62
resistors 392–3
current–voltage characteristics 409
in parallel 403–4
potential difference across 408
semiconductors 410, 478–9, 486
in series 402
resultant 65–72
force 88
parallel forces 94–5
reverse biasing 476
reverse voltage 410
reversibility principle 322, 323, 326–7, 357
rheostat 408
right-angled isosceles prism 331–2
right-hand corkscrew rule 421–2
right-hand grip rule 421–2, 433
ring circuits 456, 459
ripple tank 245, 277
rod, centre of mass 95–6
rotor 441, 451
Rumford, Count 196–7
Rumford's cooling correction 200, 203
Rutherford, Ernest 516
Rutherford's atomic model 499–501

S-pole 418, 419, 421, 422, 424, 438
safe protection 489–90
safety
electricity 385
radioactive materials 517–18, 525
symbols 562
scalars 64
power as 136
treatment of 65
scale construction 362–5
scale divisions 33, 187
scales (graphical) 35
scientific notation 5, 20
second 4
secondary cells and batteries 461–3, 464
secondary coil 442, 443–4
secondary colours 337
secondary standards 3
security technology 290
see-saw 91–2
semiconductors 410, 473–7
atomic structure 474
doping 474–5
resistors 478–9, 486
thermometer 192
sensitivity 33, 190, 191, 192
separation, and gravitational force 79
series circuits 391, 578
components 400
current in 400
with parallel circuits 406–7
potential difference in 401
resistance in 401–3
using ammeters 407
series connections 391, 393
shadows 300–3
shell atomic model 502, 510–11
short-circuits 394–5
S.I. units 2
base units 3
derived 6–7
multiple and sub-multiple 5
and standard form 8
significant figures 13, 19–20, 28, 269
rule of thumb 14, 20, 21
silvered surface 221
sinking 159, 160
slinky 241–4, 246–8
slip-rings 440, 441, 451
slope, graph 57
slow medium 326
smoke detector alarm 532–3
smooth curve 34, 46
Snell's law 270–1, 320, 321, 323, 326
sodium-24 529, 533
soft iron 424, 426, 445
solar cells 463, 464
solar eclipse 302
solar energy 134, 141, 142, 218, 451
solar water-heater 214, 221–2
solar waves 285, 286
solenoid 423, 424, 427
solid–liquid cooling curves 175
solids 172–3
characteristics of 173
and pressure 147–9
shapes of 169–70
see also metals
sound
diffraction of 276–9
ordinary 264
pitch and loudness 279
reception of 266–7
reflection of 269–70
refraction of 270–2
speed of 267–9, 270, 271–2, 472
sound waves 254
interference of 273–6
production and propagation 264–6
spark plus 375
specific heat capacity 198–9
importance of 209–10
liquids 200–2, 203–4, 209, 567
metals 202–4, 568
method of mixtures 203–4, 566, 568
specific latent heat 204
of fusion 205–6
importance of 209–10
of vaporisation 206–8
spectrum 503
continuous 285
electromagnetic 286–8
impure 334–5
pure 336
see also visible spectrum
speed
definition 105
and distance–time graphs 105–6
and kinetic energy 132–3
and power 137–8
radiation emissions 521
speed of light 267, 325, 326
speed of sound 267–9, 270, 271–2, 472
speed–time graph 107, 109, 110, 111, 118, 121
spent fuel 546
spherical lenses 344
spikes 459
spiral spring extension 36–40, 83, 88–9
split-ring 434–5
spray painting 373
spreading the error 21, 269
spring constant 38, 89
spring model of forces 173, 184
stability 100–1
stable nuclei 508, 524
stable states 512
standard form 5–6, 8, 20
standard pressure 187
standby power 486
static electric charge 371
static electricity 371, 385
stator 435, 441, 451
steam 204, 207–8, 210
steam point 187
steam turbine 545
step-down transformer 443, 453
step-up transformer 443, 453
sterilisation 534
stopwatch 21, 22–3, 268–9
straight line graph theory 57
stroking 426
strong nuclear force 508
styrofoam container 200, 202, 203, 216
sub-multiple units 4–5
submarines 161
subsonic sound 280
sums 18–19
superposition 273–5
surges 459
suspension 54
sweating 176, 210
sweep 471
switches 395, 458
electronic 483, 485, 487, 491–2
magnetic 484
main 455–6
mechanical 484
relay 485
systematic errors 29–30

telecommunications 332, 333
television waves 285, 287, 288, 291, 473
temperature 172
base units and symbols 3
body 209
change in 199
and hotness 184
and internal energy 184–5
measurement of 186–90
scales 186–8
sensitivity to 185
and sound refraction 271–2
vs pressure of gas 232–4
vs volume of gas 231–2
temporary magnets 426
tensile force 83
tension 81, 83, 92–3
terahertz rays 290
theory of cooling 44
thermal conductivity 215–16
thermal energy 135, 184–5, 213
see also heat
thermal expansion 178–9
thermionic emission 469
thermistors 478, 486
thermocouple thermometer 193
thermometers 183, 186
analogue 15
clinical 191–2
constant-volume gas 233
digital 16
electronic 189–90, 192, 478
laboratory 191
liquid-in-glass 179, 190–1, 192
physical properties 190
required features 190
thermocouple 193
weather 192–3
thermostat 179
Thomson's atomic model 497–8, 498, 499
thorium 529, 530
Three Mile Island, USA 544, 546
thrust 81, 146, 147–8
thunder 267
ticker-timer device 117–20
time
base units and symbols 3
measuring 266–7, 472
time-base voltage 471, 472
total eclipse 302, 303
total internal reflection 330–3
transformer-based filtering (TBF) 459
transformers 442–5
current and power 444
practical 445–6
principle of 443
role of 453

transistors 479
field-effect 485, 491–2
transmitted ray 323
transverse waves 243, 249
displacement 274
superposition 273
wave profile 255
travelling (progressive) waves 242, 243, 248
tree age 534
triangle method 66, 69–70
trigonometry, vector resolution 73
triple point 188
troughs 242, 243, 245, 257
true value 10, 11
truth tables 487, 488
turbines 450, 452
tweeters 275–6

ultrasound 264, 279–80
ultraviolet (UV) rays 285, 287, 289
umbra 301, 302, 303
uncertainty 10–11, 268–9
maximum 12–13, 16
range of 12
see also errors
unified atomic mass units 507
uniform acceleration 107
uniform scale 13–14, 17
analogue instruments 15
unit of electricity 460
units of measurement 2
derived 6–7
famous scientists 7, 8
fundamental quantities 3
multiple and sub-multiple 4–5
unstable nuclei 523, 524
upper fixed point 187, 189–90
upright image 315, 358, 359–61
upthrust 156–7, 158, 159, 161
uranium 516–17, 529, 534–5
nuclear fission 541
in nuclear reactors 543–4, 545
use of knowledge (UK) 31

vacuum, sound travelling in a 265–6
vacuum flask 221
vacuum-tube devices 469
valence electrons 473–5
Van de Graaff machine 380, 381
vapour pressure 177
vapours *see* gases
variable capacitors 380
variables 557
controlling 44
dependent and independent 28, 29, 33, 34, 46
vector equation 67
vectors 64
angle between 67
momentum 112, 113
parallel 70–2
parallelogram method 67–70
rectangle method 72–3
resolution into two components 72–3, 128
treatment of 65–7
velocity 106, 111
velocity 6
definition 105
and displacement–time graph 106–7
positive and negative 111
as vector quantity 64, 72–3
velocity–time graph 107, 108–10, 111, 124
ventilation, buildings 222–3
vernier caliper 12, 14, 30
vibrating systems 173, 184, 197
vibrations *see* oscillations
virtual focus 345, 347
virtual image 310, 313, 358, 359–61
location of 354–5
visible rays 285, 289
diffraction and interference 291
visible spectrum 287, 289, 333, 503
recombining colours of 336–7
volatile liquids 175, 208–9
volt 390, 391

voltage
forward and reverse 410
input 443, 487–8
low 459
time-base 471, 472
see also potential difference
voltage reversal protection 485–6
Volta's cells 386, 460
voltmeters 193, 390
oscilloscope as 472
in a series circuit 401
using 408
volume 6
vs pressure of gas 227–30
vs temperature of gas 231–2

warning signs 385, 517, 562
wasted energy 139, 140
water
anomalous expansion 179–80
change of state 204
convection in 217
heavy water 544
molecule 169
specific heat capacity 202, 566
specific latent heat of vaporisation 207–8
thermal conductivity 216
thermal properties 209–10
water circuit 400
water heater control 488, 490
watt (W) 137
wave parameters 252
wave–particle duality 296, 297
wave-probability model 504, 511
wave profile 245, 256
wave speed 258–9
wave-trains 276
waveforms 472–3
wavefronts 245, 246, 257, 296
and rays 258
shape of 258
sound 272
wavelength 256, 257
and diffraction 277–8
electromagnetic radiation 285, 286
electromagnetic spectrum 287
emission 502–3
ultrasound 280
waves 240–1
cycle of 242, 253, 254–5
definition 241
diffraction 276–7
matter 243, 249–50
production 241–3
in a ripple tank 245
travelling through a medium 246–9
see also electromagnetic waves; longitudinal waves; transverse waves
weather balloon 235
weather thermometer 192–3
weight 78
S.I. unit of 80
vector 64
white light
dispersion 333–5
pure spectrum from 336
from spectral colours 336–7
white surfaces 222
wind farms 451
windmills 142
Windscale, UK 546
wire, resistance of 556–62
wiring circuits 396
wood 141, 161
work 127, 128–9, 130, 564
and heat 197
and power 136

X-deflection plates 470, 471
X-ray diffraction 291
X-ray imaging 280, 288
X-rays 285–6, 287, 288

Y-deflection plates 470, 471–2
Young, Thomas 296
slits experiment 298–9

zero errors 18, 29–30
zeroing instruments 30

Physics for CSEC® Examinations

3rd Edition for the 2013 syllabus

Physics for CSEC® Examinations is part of a well-established series of books aimed at students preparing for their CSEC Science studies. Rejuvenated in a third edition, **Physics for CSEC® Examinations** features comprehensive, systematic coverage of the latest CSEC syllabus (2013). Written by an expert team of science educators, this revised edition benefits from a new, clear and accessible design and the most up to date scientific information.

Key features of the CSEC Science series:

- Intuitive and easy-to-follow format makes it simple to study a whole topic, or to find answers to specific problems
- Regular consolidation (in-text questions and exam preparation) checks understanding and reinforces learning
- New group-work feature tests students' investigative and problem-solving skills and demonstrates real-world applications of key syllabus points
- Practical activities and experiments throughout the text encourage hands-on learning
- Dedicated School-Based Assessment section gives step-by-step tips to maximise success in the CSEC coursework.

Alec Farley BA (Physics), MS (Physics), Diploma in Advanced Studies in Science Education, is a former Chief Examiner in CXC Integrated Science. He has served as Education Officer, Science and Mathematics within the Ministry of Education, Guyana and has been the head of both Mathematics and Physics departments throughout the region.

Clarence Trotz MA (Cantab), Cert Ed. was a member of the panel which formulated the first CSEC physics syllabus, and went on to become Chief Examiner. He has also served in the Ministry of Education of Guyana as Co-ordinator of Science and Mathematics Education.

Series Editor: Dr. Mike Taylor (MA., PhD., CChem., FRSC.) has been actively involved in education and teaching for many years. He has considerable experience of teacher training, has examined science at 'O' and 'A' levels all over the world, and has taught chemistry at School and University levels for over forty years.

Also available in the CSEC Science series:



For more innovative content, log on to find
FREE Online Teacher's Resources
www.macmillan-caribbean.com/resources

e book

CSEC® is a registered trade mark of the Caribbean Examinations Council (CXC). Physics for CSEC® Examinations is an independent publication and has not been authorized, sponsored, or otherwise approved by CXC.

MACMILLAN
www.macmillan-caribbean.com

CONNECT WITH US:

Find us on Facebook
www.facebook.com/macmillancaribbean

Find us on Twitter
@MacCaribbean

ISBN 978-0-230-43884-2